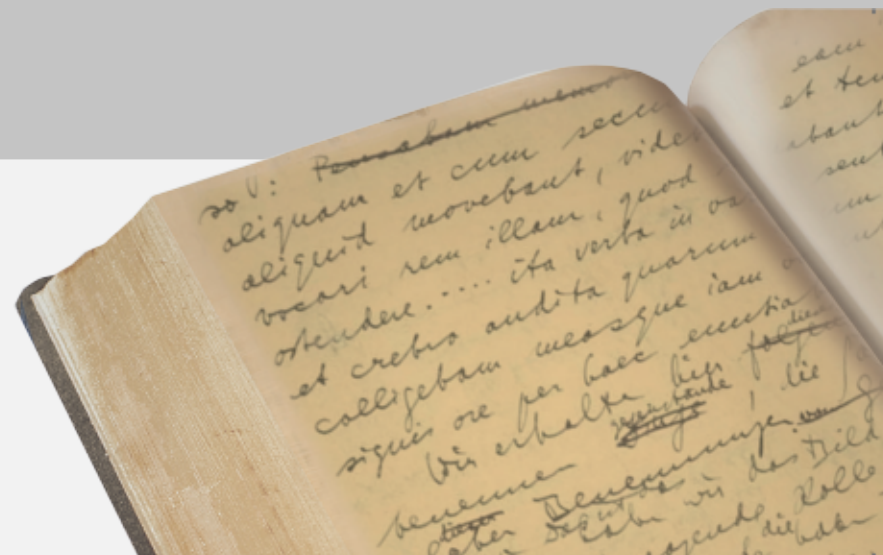


# From Records to Graphs: Linked Data and Libraries

*Prof. Dr. Stefan Gradmann (KU Leuven)*

Universidad Carlos III de Madrid, 11/07/2013



## Data about Data: Objects and Assertions

- Object/Data Models, Assertion Models and Pragmatics

## From Containers to Context, from Records to Graphs, from Catalogs to Research Data:

- Container → Content → Context: Libraries approaching the End of the Gutenberg Galaxis and the emerging Linked Data Web
- The Europeana Data Model (EDM) in this context
- EDM (and RDF) enabling Research with content based and context driven services
- Required Cultural Changes: terms/thinking to get rid of

## Hands on: Semantic Annotation using Pundit

# Metadata. Meta Data?

## What is this Talk About?

# Not about ...

- ... cataloguing practice
- ... manual vs. automated metadata generation
- ... named entity detection
- ... metadata licensing issues
- ... metadata quality issues
- ...



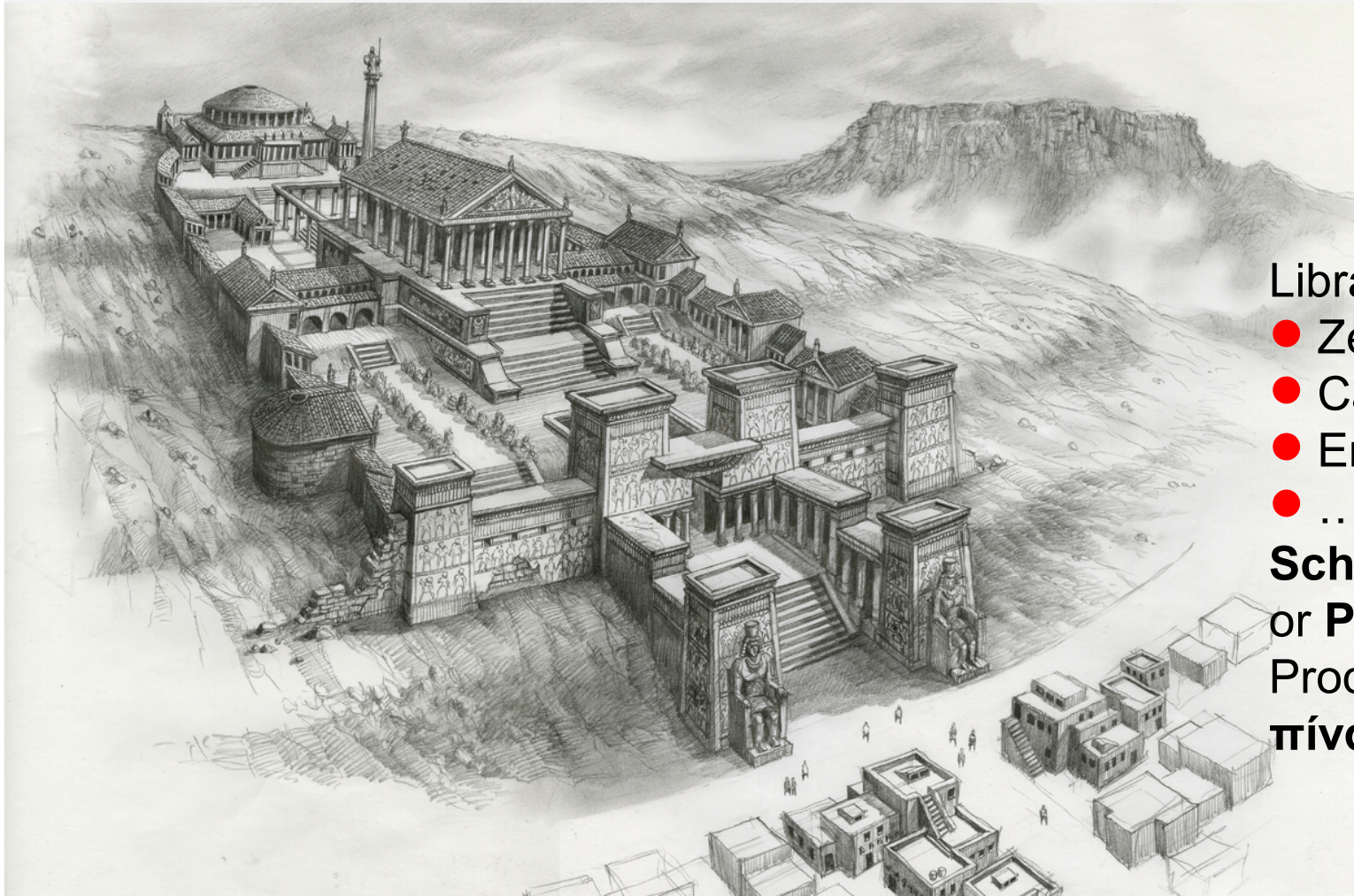
# But rather about ...

- Data about data?
- Statements on things in the world
  - Like in  
<http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DR.html#DREMB>
  - Or in  
[http://primocat.bl.uk/F?func=direct&local\\_base=PRIMO&doc\\_number=012233614&format=001&con\\_lng=prm](http://primocat.bl.uk/F?func=direct&local_base=PRIMO&doc_number=012233614&format=001&con_lng=prm)
- Having
  - A subject: aboutness (far from being obvious!)
  - A grammar: a data / object model (mostly implicit)
    - <http://people.umass.edu/phil335-klement-2/tlp/tlp.html>
  - A Syntax: coding and serialisation
  - A pragmatic context: an implied audience
- → This talk is restricted to bibliographical metadata in libraries

# Container → Content → Context

## Libraries approaching the end of the Gutenberg Galaxis

# Long before the Parenthesis: Alexandria



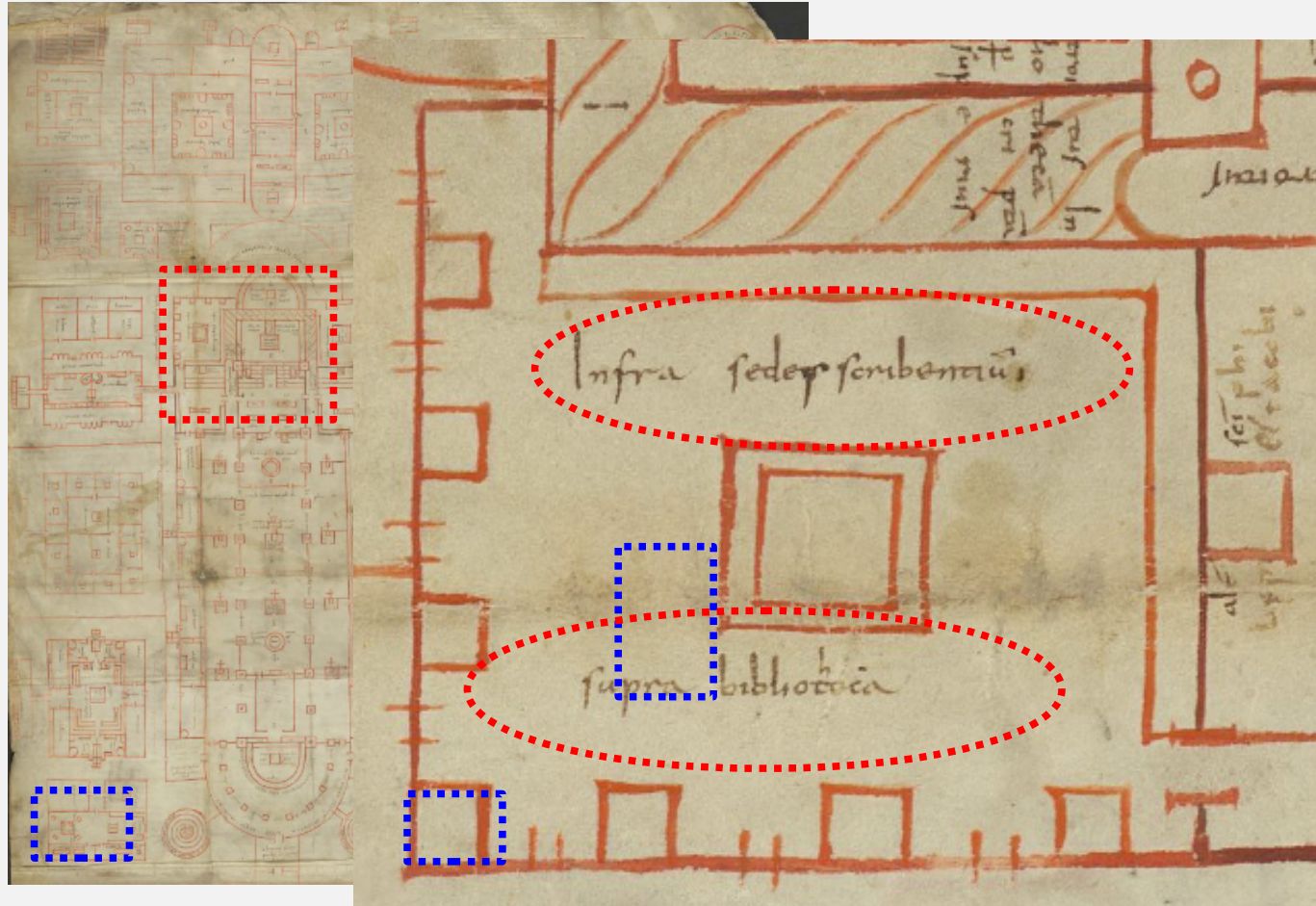
Librarians:

- Zenodotus
- Callimachus
- Erathosthenes
- ...

**Scholars** and /  
or **Poets**  
Producing  
**πίνακες**



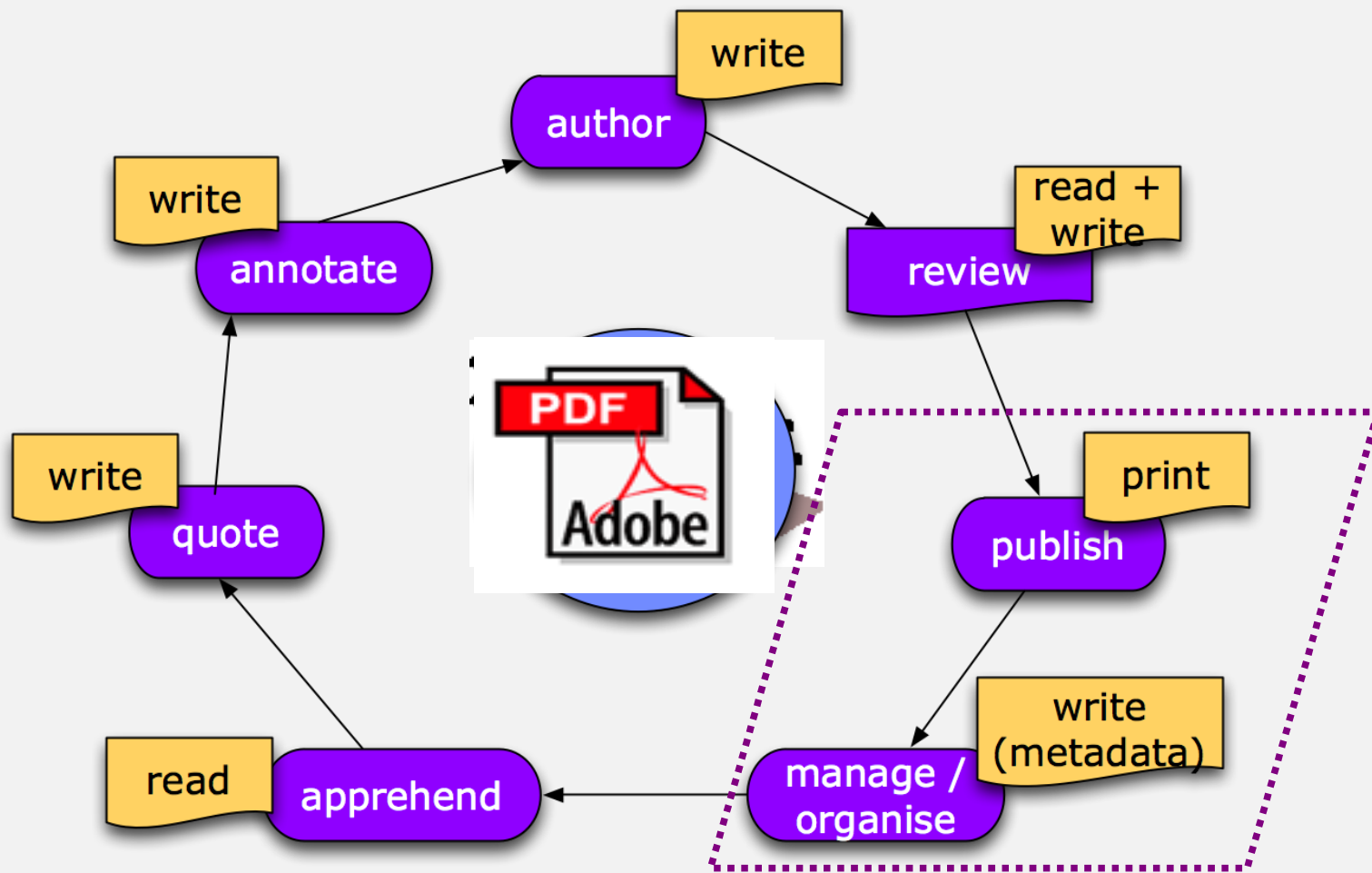
# Before the Parenthesis: St. Gall



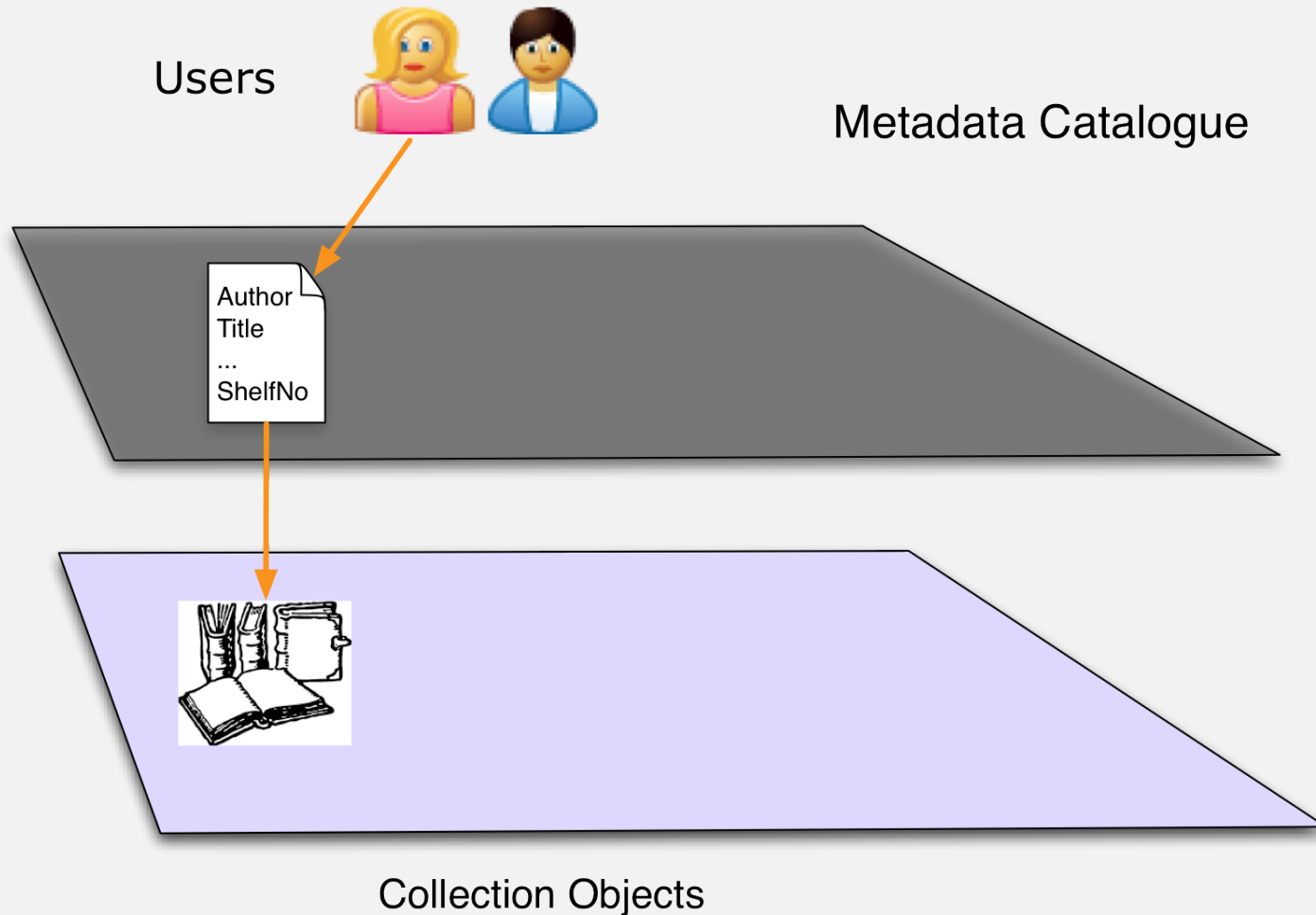
# The Gutenberg Parenthesis Opens ...

## Dissociation of container and content in the print paradigm

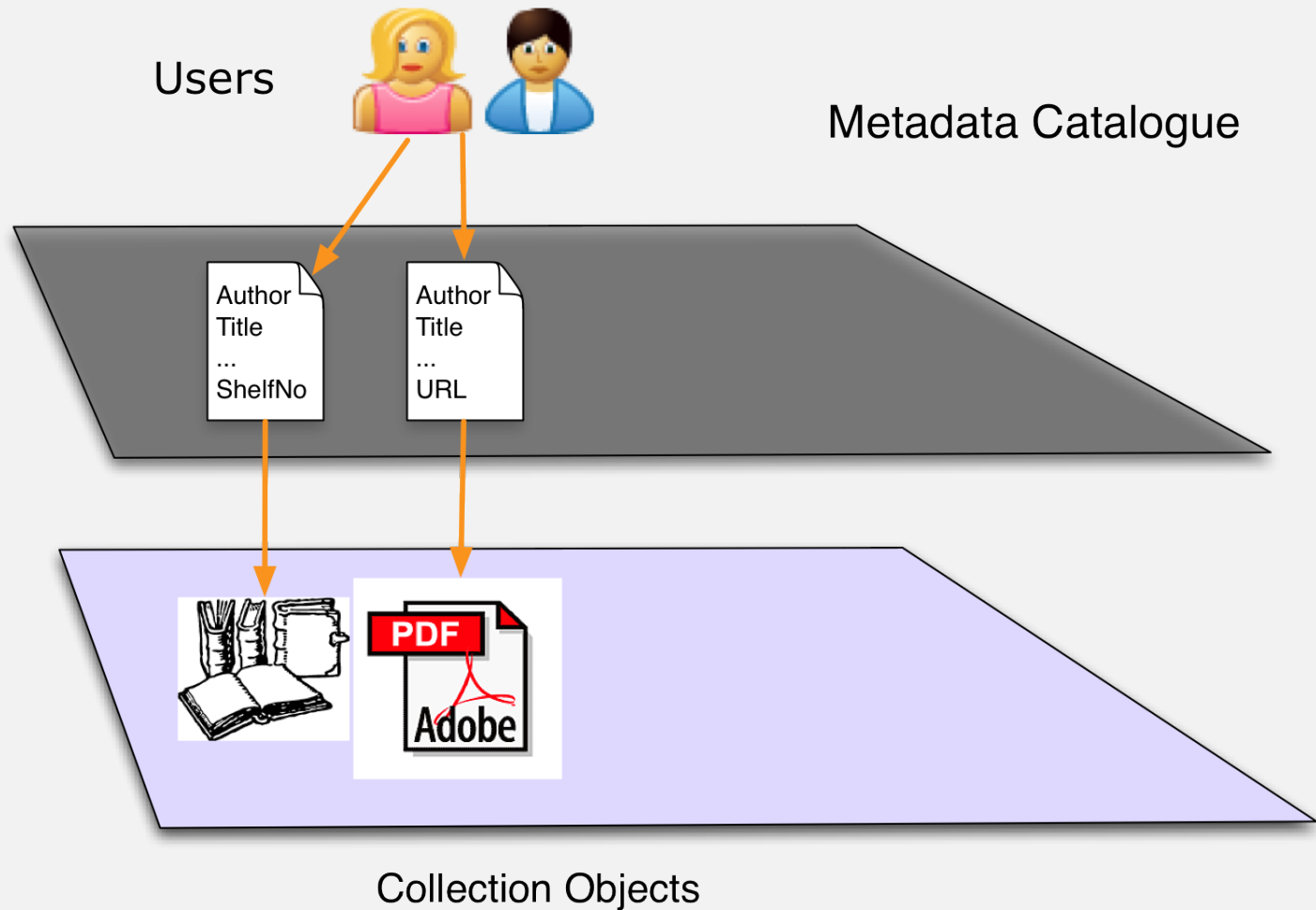
# Dissociation of Roles in the Gutenberg galaxy



# Catalogue Based Library Functional Axioms (1)

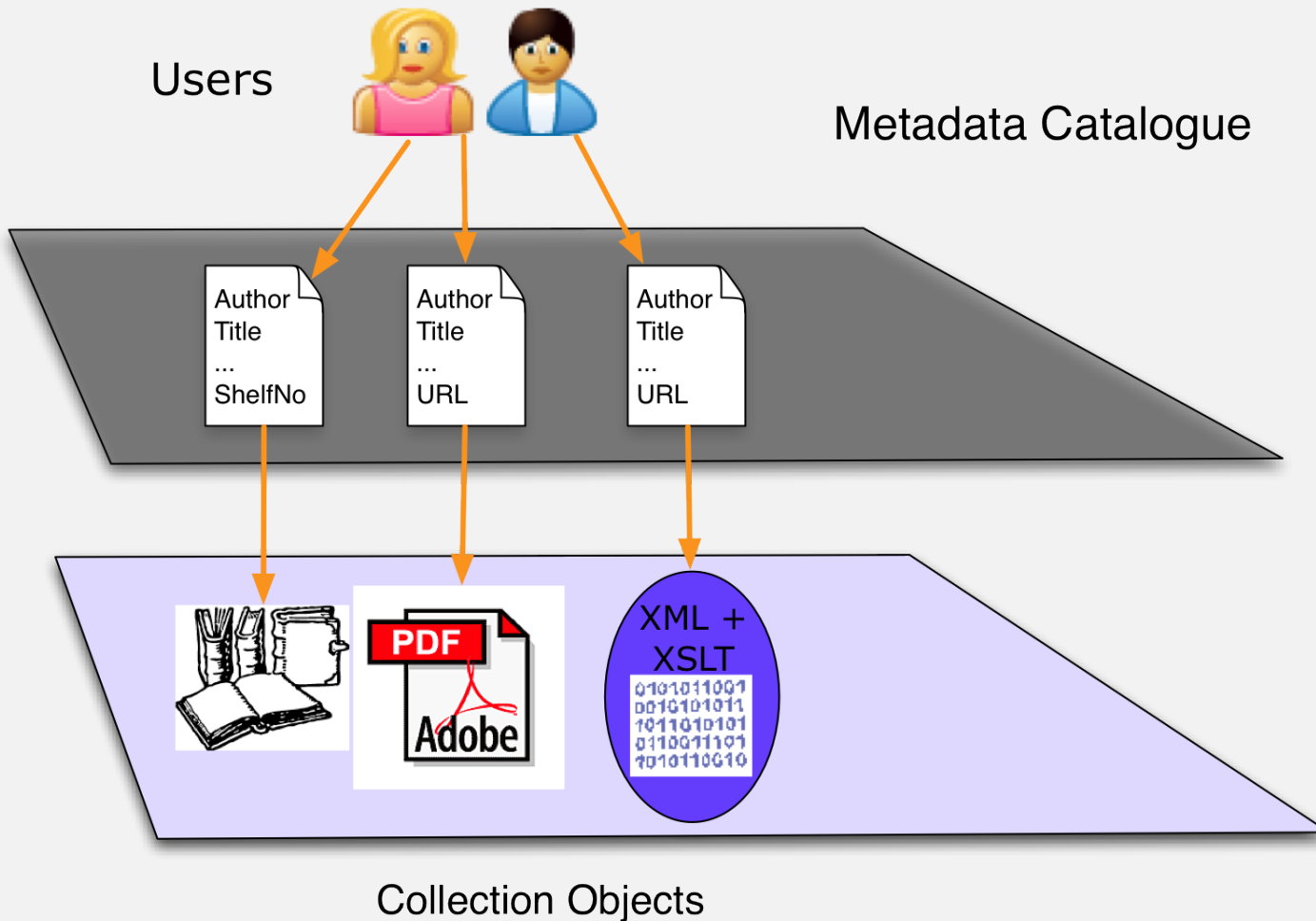


# Catalogue Based Library Functional Axioms (2)

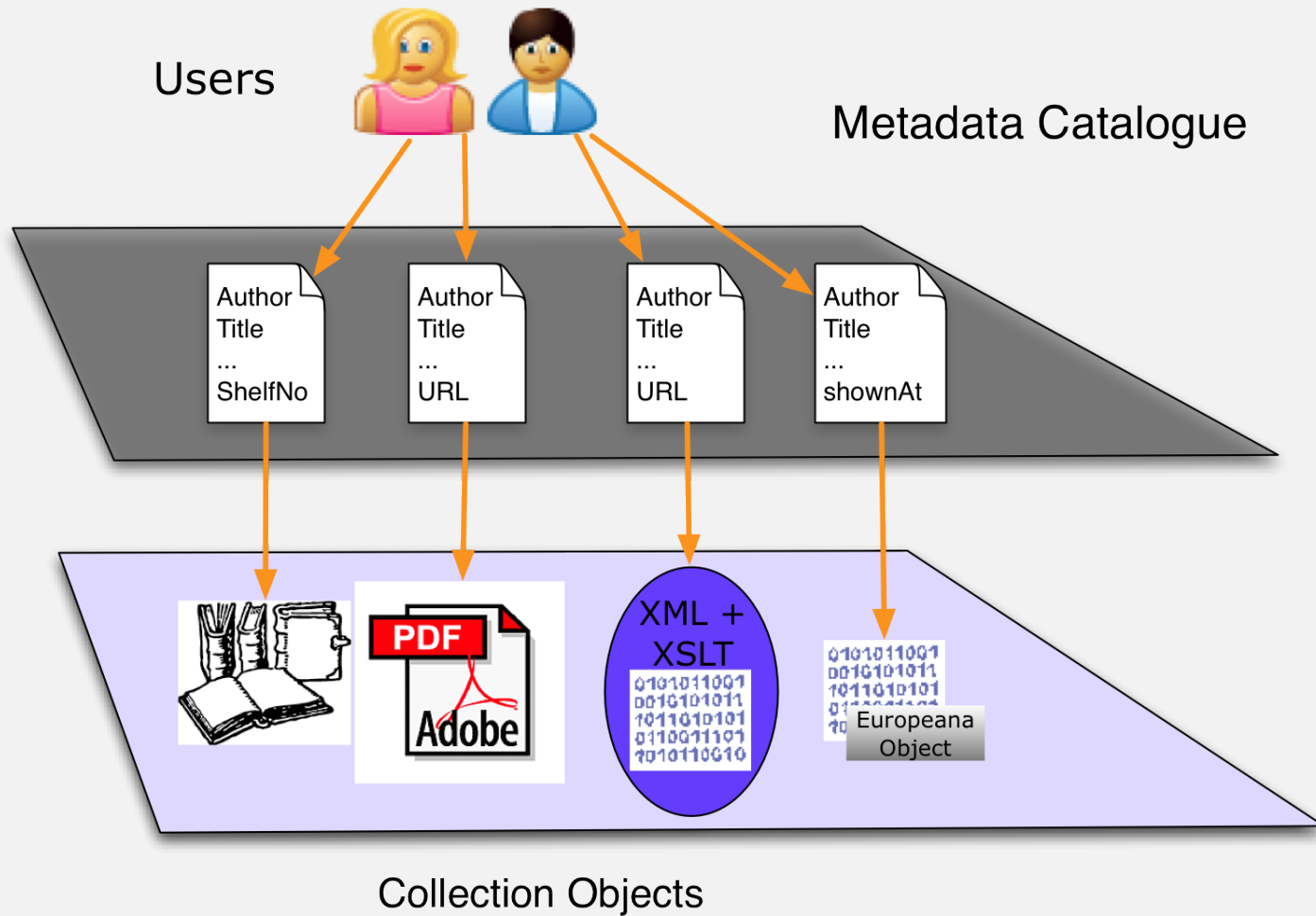




# Catalogue Based Library Functional Axioms (3)



# Catalogue Based Library Functional Axioms (K) LEUVEN



- **Mediating access** to information objects via **catalogues**
- **Mediating links** as pointers from metadata to objects
- Objects are part of a library **collection**
  - An object to be used within a library typically is part of this library's collection
- Internal processing logic: focus on
  - objects as information **containers**,
  - not so much on the **content** of these containers
  - and accordingly **cataloguing** is focussed on **container attributes**
- Functional macro-primitives are **ingestion, storage, description** and **retrieval** of information containers

## ... and examples: metadata records

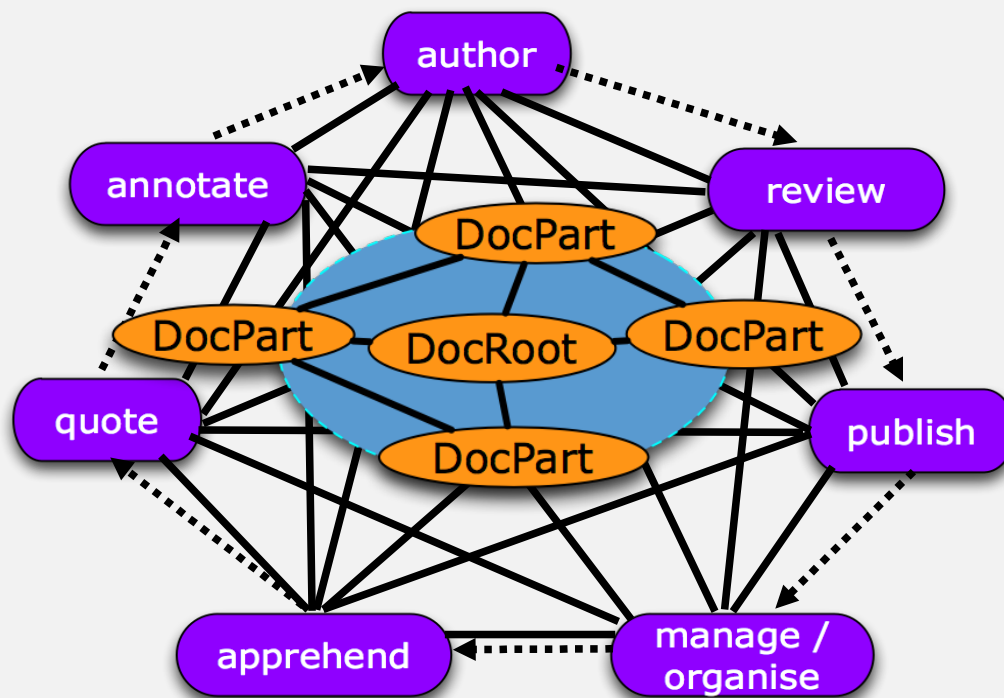
- <http://lccn.loc.gov/72083804>
- <http://lccn.loc.gov/00711195>
- <http://catalogue.bnf.fr/ark:/12148/cb34605779c>
- <http://d-nb.info/574200282>

# ... and closes again

## The end of the print paradigm

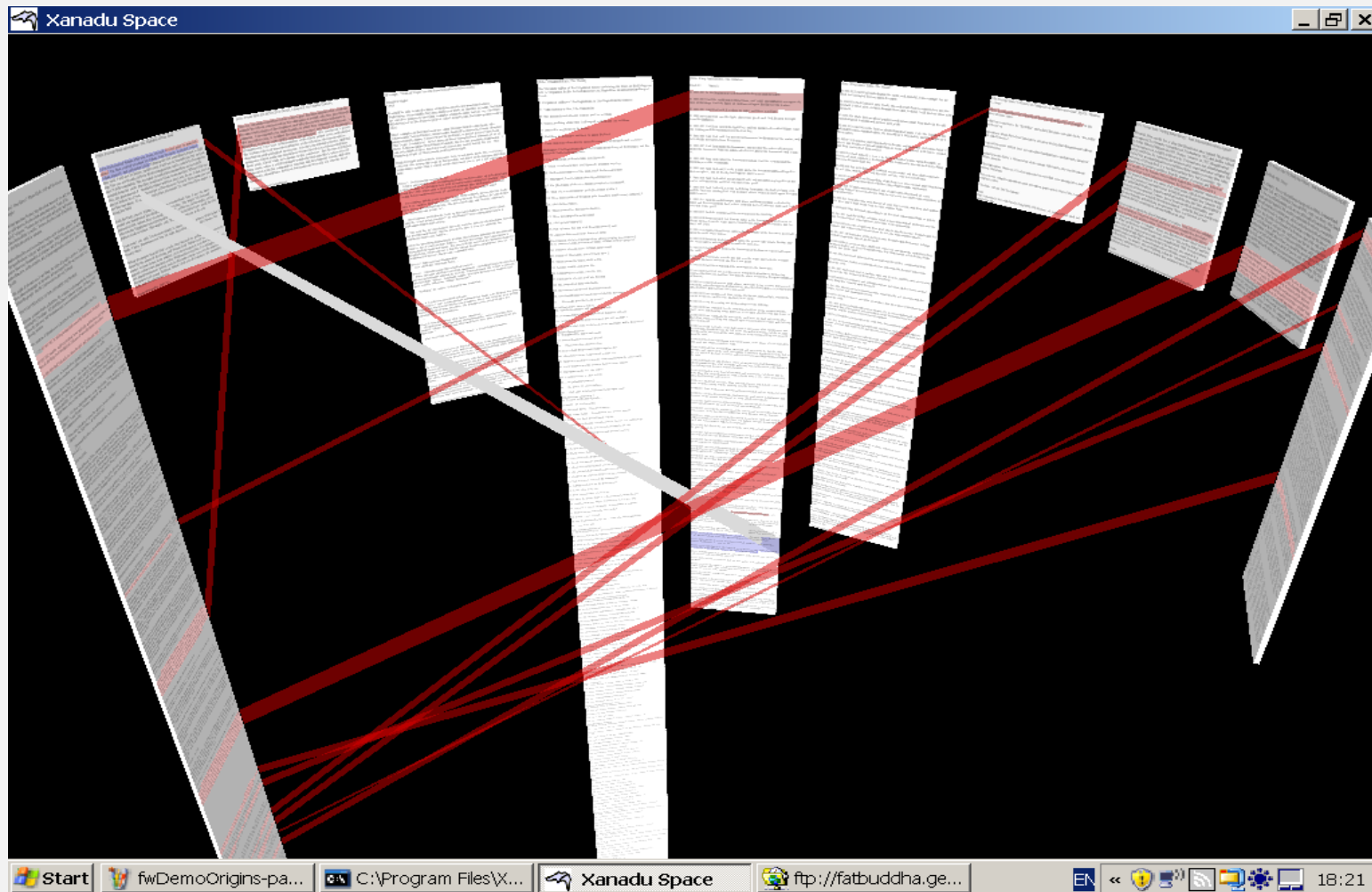
# Web Based Scholarly Working Continuum ... ... a triple paradigm shift: Beyond Documents

KU LEUVEN

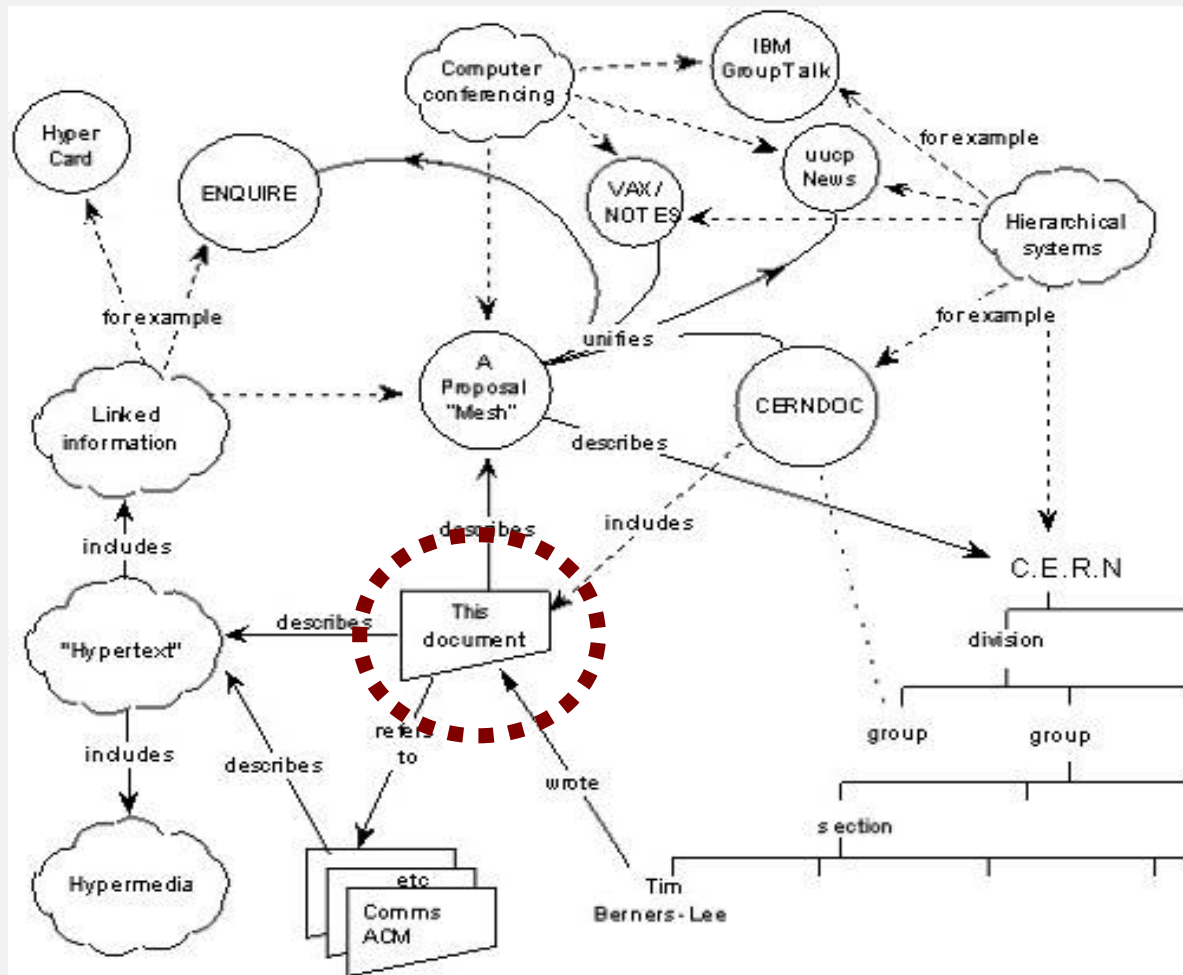


- Decreasing functional determination by traditional cultural techniques
- Disintegration of the linear / circular functional paradigm
- Erosion of the monolithic document notion in hypertext paradigms

# Ted Nelson's Xanadu: radicalised Hypertext ...



# The Web of Documents

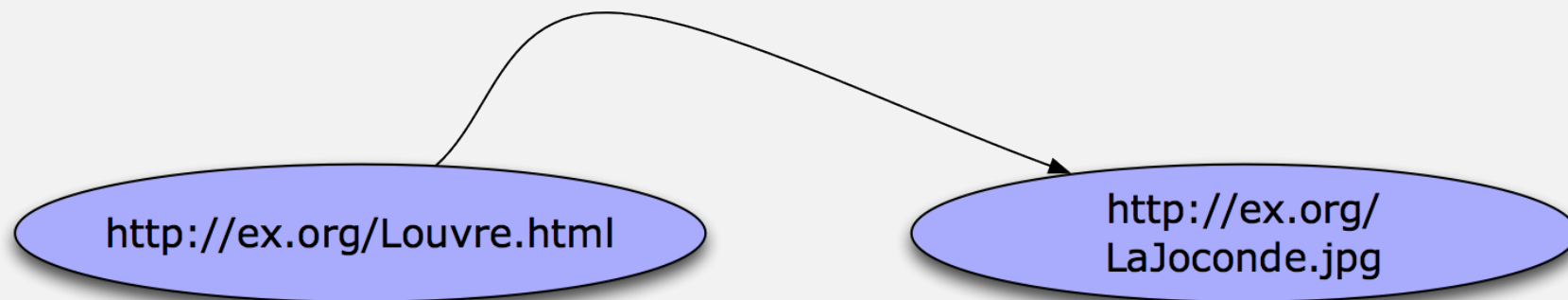


Information Management:  
A Proposal  
(TBL, 1989)

... twice  
extended:  
• in syntax  
• in scope



# Resources and Links in the Document Web

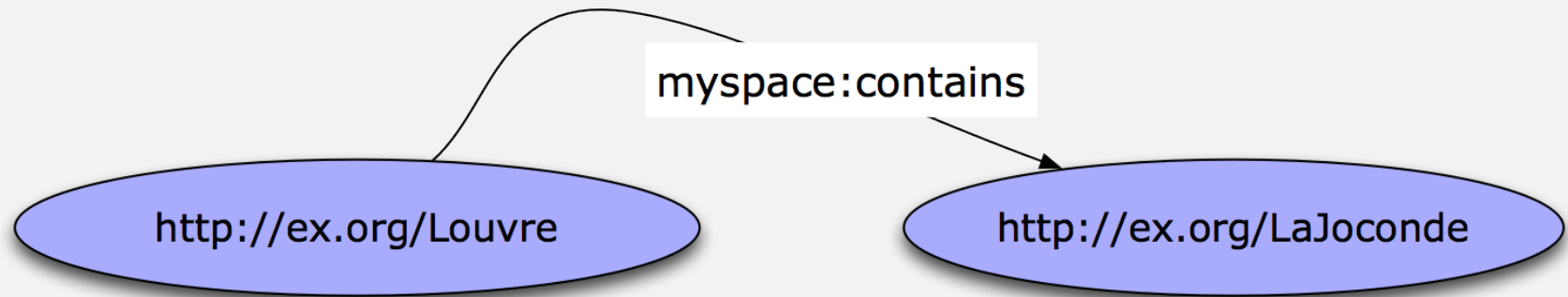


```
<a href="http://ex.org/LaJoconde.jpg">b</a>
```

- We have HTTP URIs to identify resources and links between them – but we are missing a few things!
- What kinds of resources are 'Louvre.html' and 'LaJoconde.jpg'?
  - A machine cannot tell.
  - Humans can: we recognize implied context!
- How exactly do they relate to each other?
  - A machine cannot tell.
  - Humans can: again we recognize implied context!

# Syntactically Extending the Document Web (1)

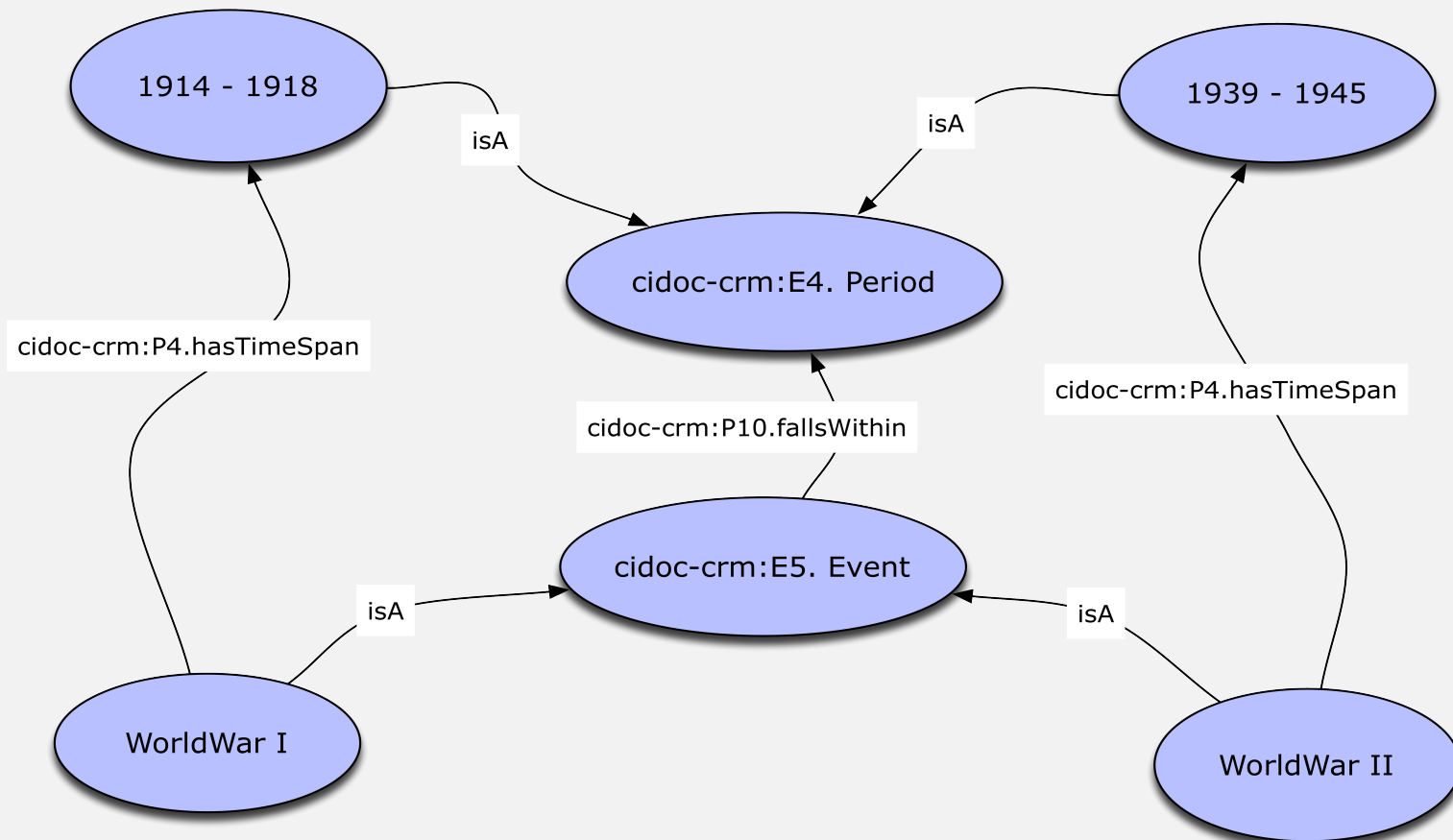
- We add a syntax for making statements on resources: RDF triples



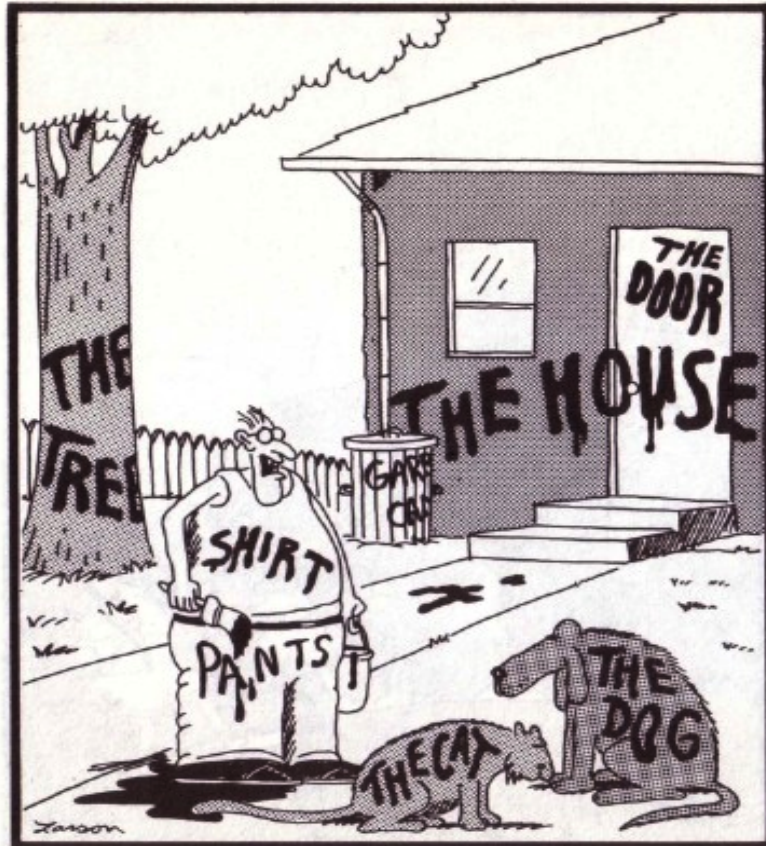
- We add a schema language (RDFS) with elements such as
  - classes (chair' as instance of chairs),
  - hierarchies of classes and properties (chairs are a subclass of furniture, 'teaches' is a sub-property of 'communicates')
  - inheritance (communication based on language → teaching also is)
  - support for basic inferencing, deterministic logical operations

# Syntactically Extending the Document Web: RDF (2)

- And thus are able to establish structures in triple aggregations resulting in lightweight domain ontologies:



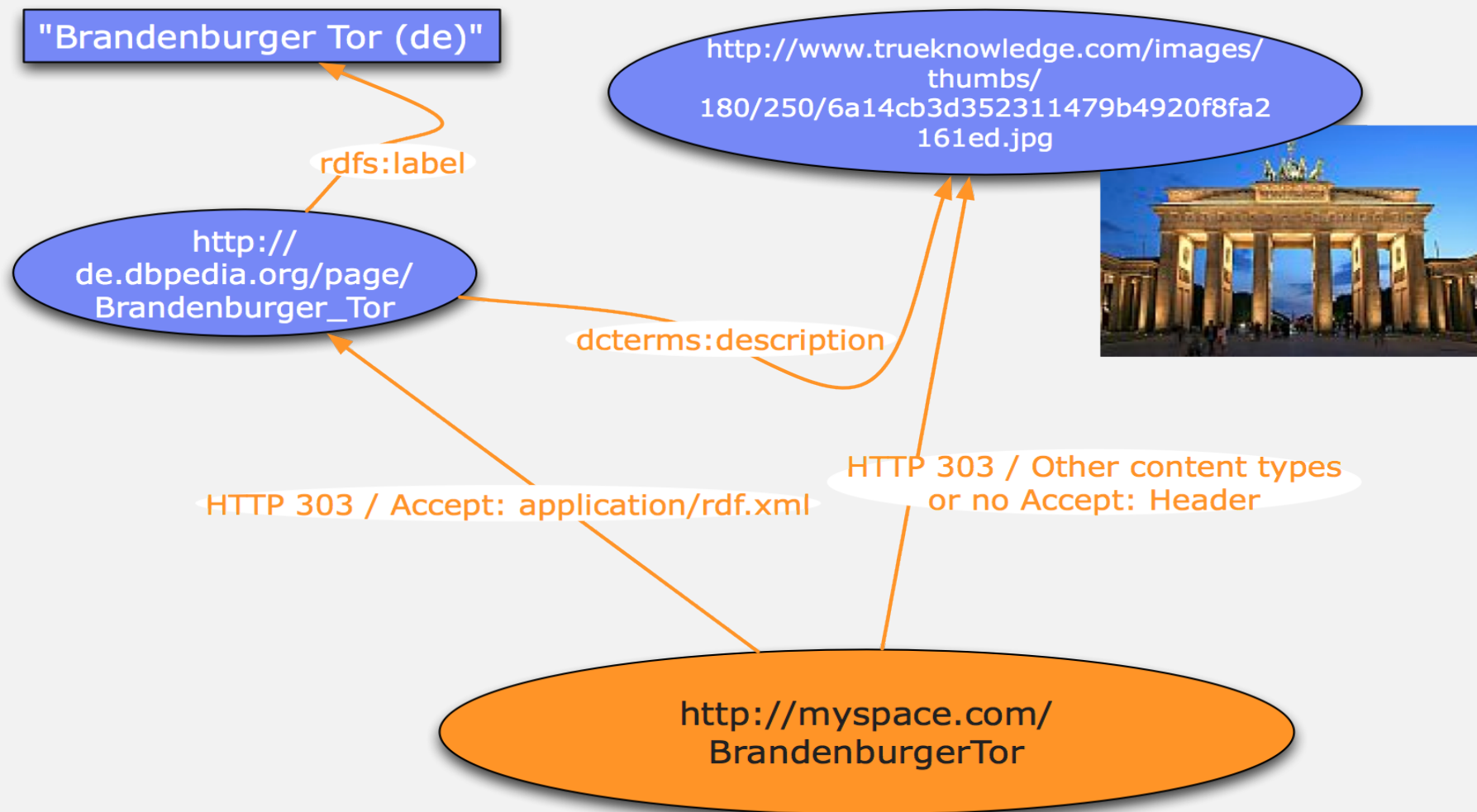
# Extending the Web in Scope: The Web of Things ... (slightly Mistaken)



What's **wrong**  
with this picture?

Taken from Ronald Carpentier's  
Blog at  
<http://carpentier.wordpress.com/2007/08/08/1-2-3/>

# ... and the Way we extend the Web in scope to make it a 'Web of Things'



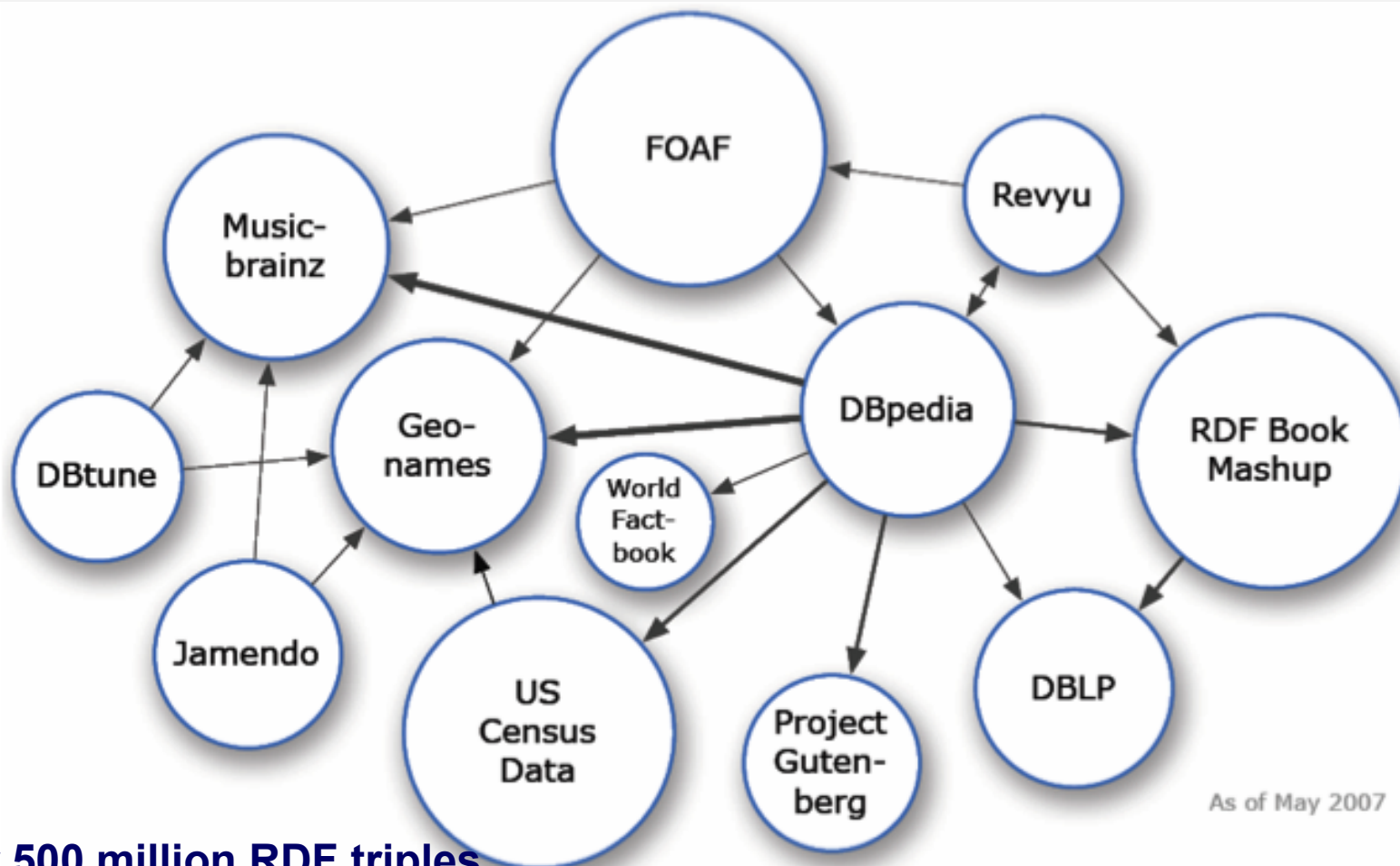
## Linked Data essentials

1. Use URIs      Standard **Identifiers**
  2. Use HTTP URIs      Standard **Pointers**
  3. Serve useful information using SPARQL, RDF standards
  4. Mention URIs of related objects      Standards for **Queries** and **Statements**
- Link** to Context

Copyright © 2008 W3C (MIT, ERCIM, Keio)

[http://www.w3.org/2008/Talks/0617-lod-tbl/#\(4\)](http://www.w3.org/2008/Talks/0617-lod-tbl/#(4))

# A few Bubbles: 5/2007



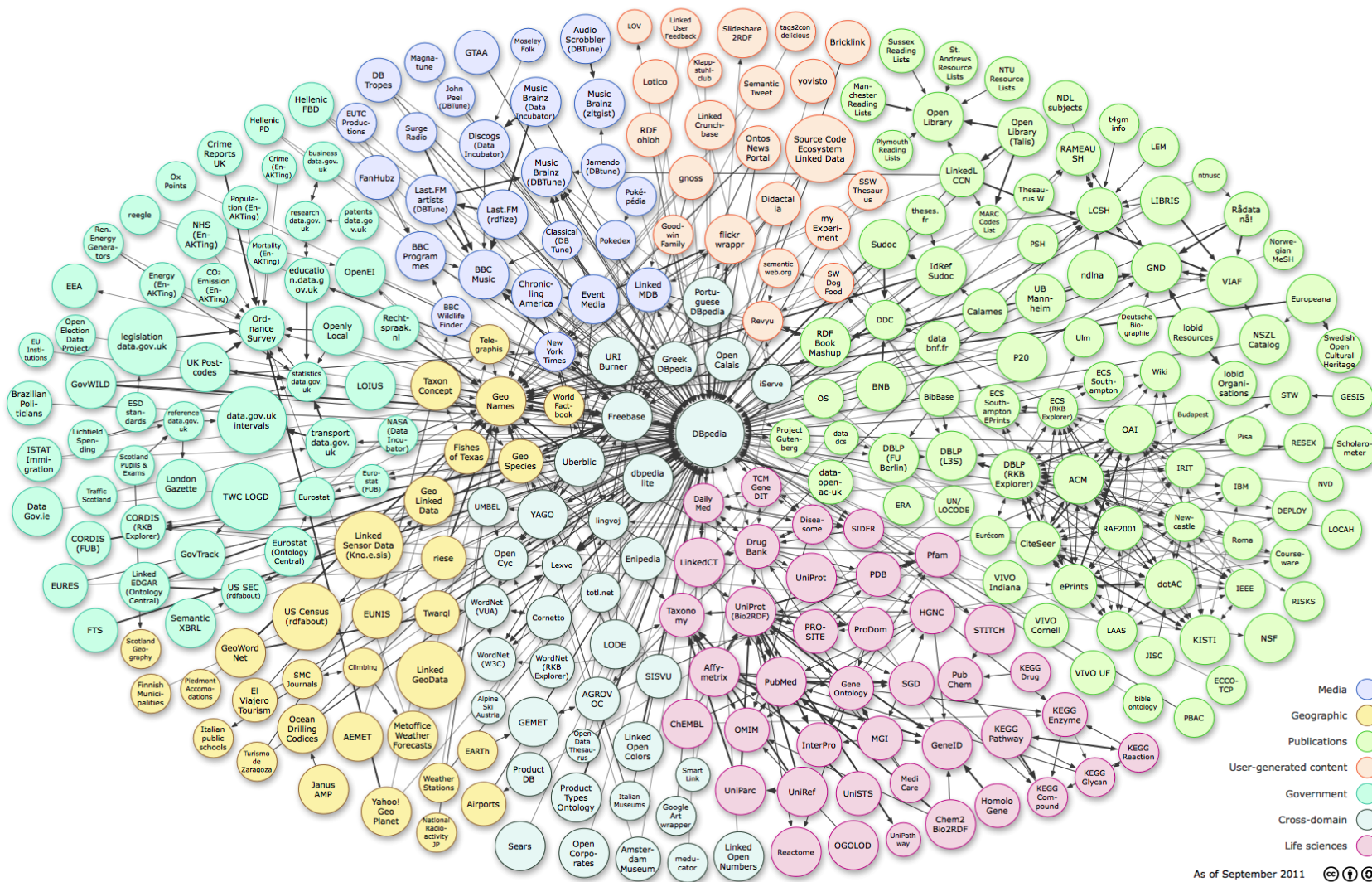
As of May 2007

- Over 500 million RDF triples
- Around 120,000 RDF links between data sources

© Richard Cyganiak

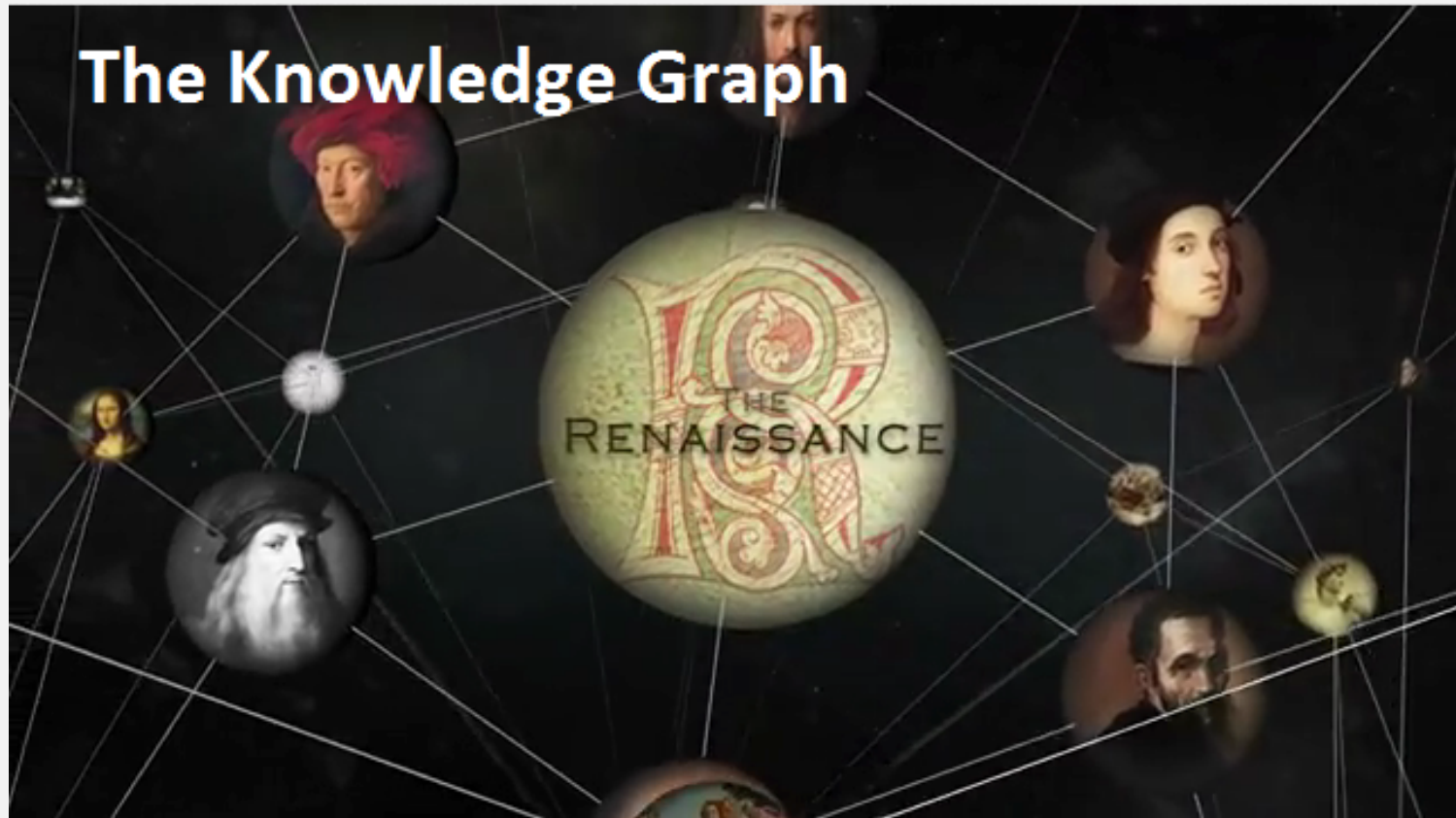


# And a lot of Bubbles as of last Year



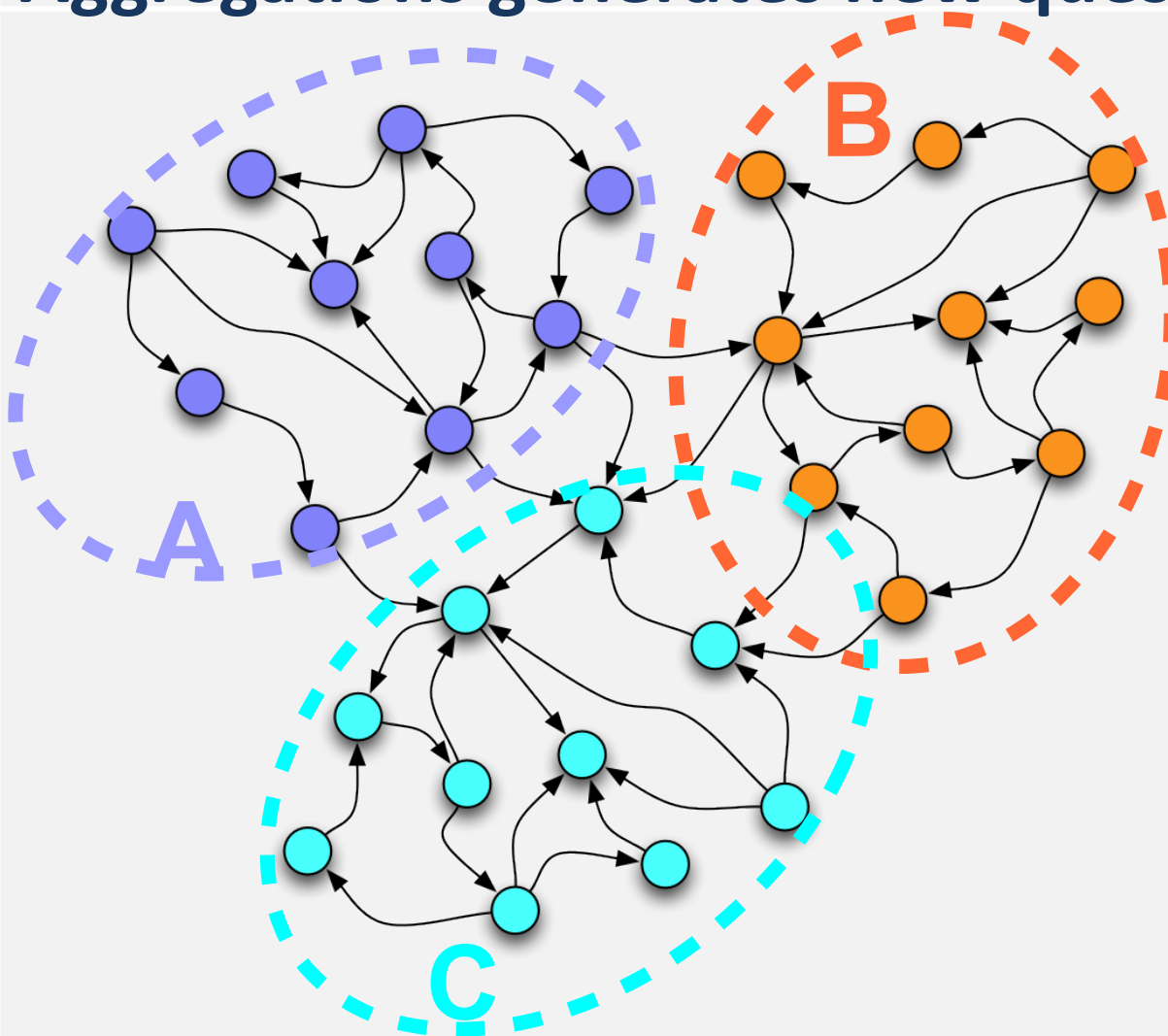


# Google entering the Floor



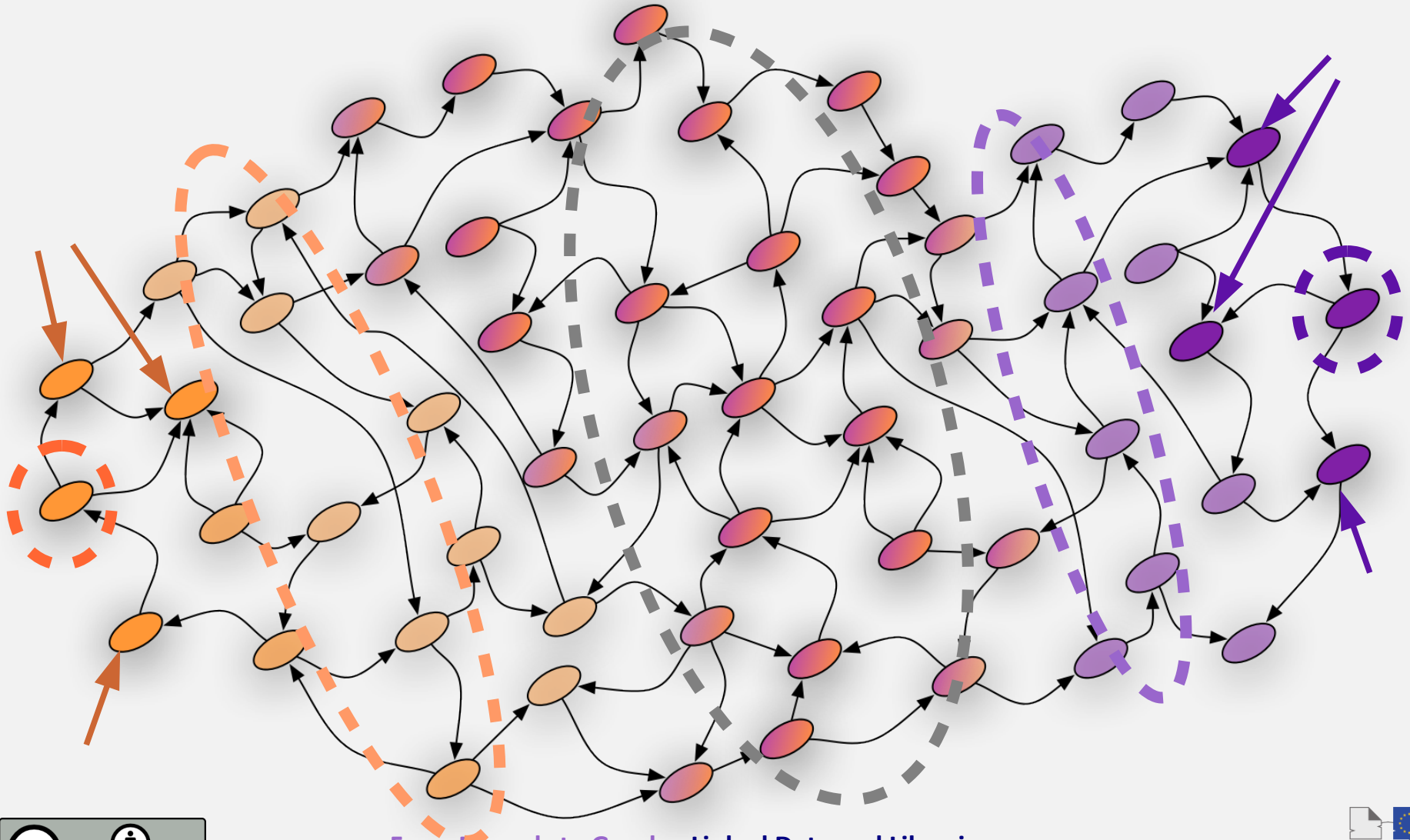
# Modelling Object Representations as RDF

## Aggregations generates new questions ...



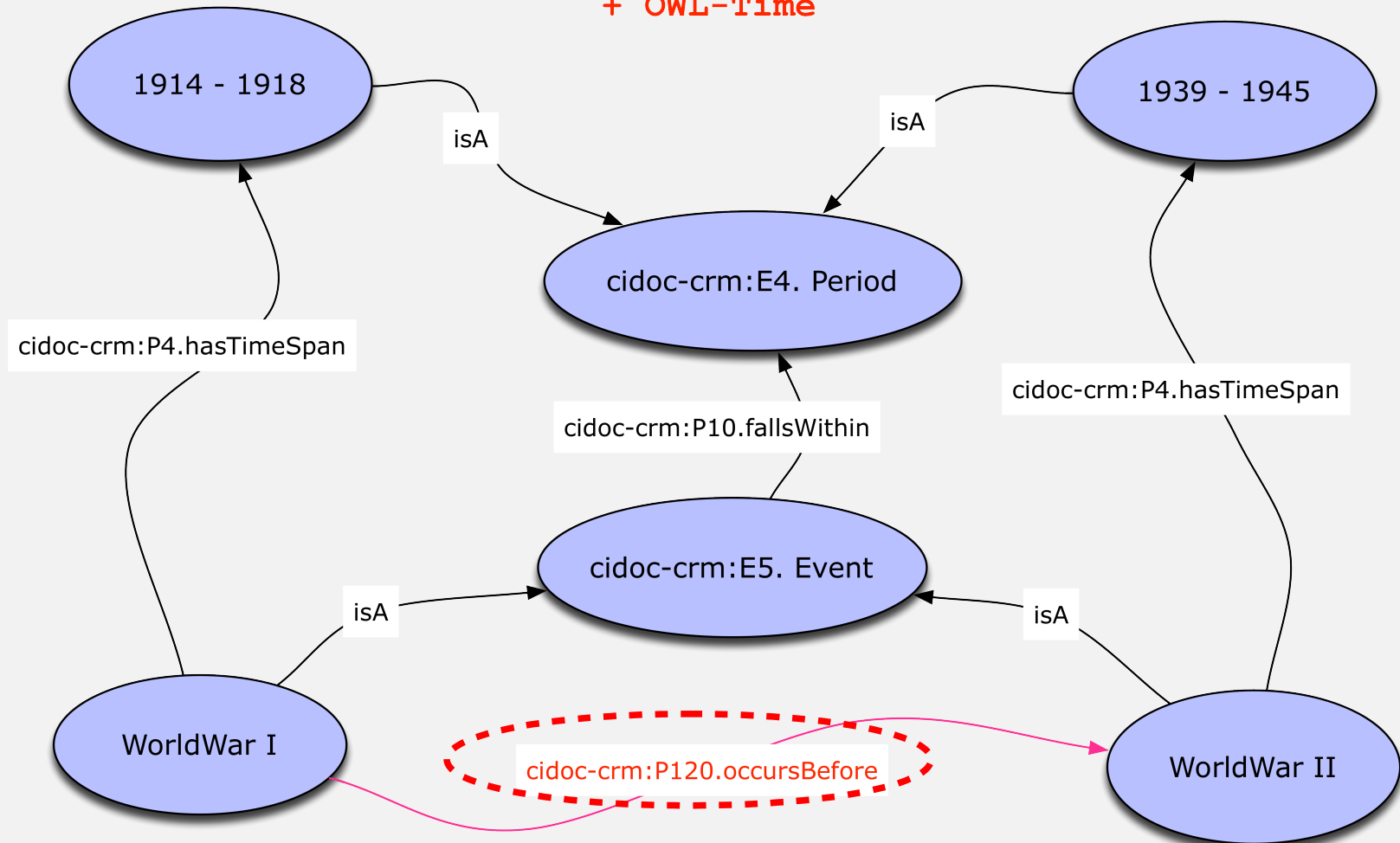
- Where do resource aggregations 'start'? Where do they 'end'?
  - And what constitutes document boundaries??
  - And which node was connected to which one at a given time???
- Provenance,  
Versioning,  
Authorisation: Named  
Graphs

# Aggregations and Context: Calculating Closeness



# ... and new opportunities: Triple Sets and 'Reasoning'

+ OWL-Time



# ... based on 'Documents' as Aggregations of RDF-Triples (1)

Assertion

NG\_000007.  
3:g.70628G>A

has  
frequency

0.25%

Condition

Sardinian

Provenance

Giardine et al

# 'Documents' as Aggregations of RDF-Triples (2)

```
<nanopublication id="0">
```

```
<assertion>  
  <subject>NG_000007.3:g.70628G>A</subject>  
  <predicate>has variant frequency</predicate>  
  <object>0.25%</object>  
</assertion>
```

```
<condition>Sardinian</condition>
```

```
<provenance>  
  <dateofcreation>March 24, 2011</dateofcreation>  
  <lastedit>March 24, 2011</lastedit>  
  <evidenceType>empirical</evidenceType>  
  <authorID>Giardine et. al.</authorID>  
  <curatorID>unresolved</curatorID>  
  <registrantID>Mons et. al.</registrantID>  
  <PMID>6695908</PMID>  
  <PMID>1428944</PMID>  
  <PMID>1610915</PMID>  
  <DOI>http://dx.doi.org/10.1038/ng.785</DOI>  
  <linkout>http://globin.bx.psu.edu/cgi-bin/hbvar/query_vars3?  
mode=output&display_format=page&i=239</linkout>  
  <linkout>http://phencode.bx.psu.edu/cgi-bin/phencode/phencode?  
build=hg18&id=HbVar.239</linkout>  
</provenance>
```

```
<nanopublication id="0">
```

# The use of Inferences

KU LEUVEN

van Haagen HHHBM, 't Hoen PAC, Botelho Bovo A, de Morrée A, van Mulligen EM, et al. (2009) Novel Protein-Protein Interactions Inferred from Literature Context. PLoS ONE 4(11): e7894. doi:10.1371/journal.pone.0007894 / Example provided by Jan Velterop

**PLoS one**  
accelerating the publication of peer-reviewed science

Login | Create Account | Feedback

Search articles... **GO** Advanced Search

Browse RSS

Home Browse Articles About For Readers For Authors and Reviewers Journals Hubs PLoS.org

**RESEARCH ARTICLE** OPEN ACCESS

**Novel Protein-Protein Interactions Inferred from Literature Context**

Article Metrics Related Content Comments: 0

**Herman H. H. B. M. van Haagen<sup>1\*</sup>, Peter A. C. 't Hoen<sup>1</sup>, Alessandro Botelho Bovo<sup>2</sup>, Antoine de Morrée<sup>1</sup>, Erik M. van Mulligen<sup>1</sup>, Christine Chichester<sup>1</sup>, Jan A. Kors<sup>1</sup>, Johan T. den Dunnen<sup>1</sup>, Gert-Jan B. van Ommen<sup>1</sup>, Silvère M. van der Maarel<sup>1</sup>, Vinícius Medina Kern<sup>2</sup>, Barend Mons<sup>1</sup>, Martijn J. Schuemie<sup>1</sup>**

**1** Biosemantics Association, Department of Human Genetics, Leiden University Medical Center, Leiden, and Department of Medical Informatics, Erasmus University Medical Center, Rotterdam, The Netherlands, **2** Post-Graduate Program in Knowledge Engineering and Management (EGC), Federal University of Santa Catarina (UFSC), Florianópolis, Brazil

**Abstract** [Top](#)

We have developed a method that predicts Protein-Protein Interactions (PPIs) based on the similarity of the context in which proteins appear in literature. This method outperforms previously developed PPI prediction algorithms that rely on the conjunction of two protein names in MEDLINE abstracts. We show

To add a note, highlight some text. [Hide notes](#)

[Make a general comment](#)

**Jump to**  
[Abstract](#)  
[Introduction](#)  
[Results](#)  
[Discussion](#)  
[Methods](#)  
[Supporting Information](#)  
[Author Contributions](#)  
[References](#)

**Download:** [PDF](#) | [Citation](#) | [XML](#)  
[Print article](#)

**Metrics** ⓘ

**Total Article Views:** **673**


**Average Rating** [\(0 User Ratings\)](#)  
☆☆☆☆ [See all categories](#)  
[Rate This Article](#)

**Related Content**

**Related Subject Categories**  
[Molecular Biology](#), [Computer Science](#), [Computational Biology](#)

**Related Articles on the Web**  
[Google Scholar](#)  
[PubMed](#)

**Share this Article** ⓘ

Bookmark:      

[Email this article](#)





# Data = Publication

- Distinction data vs. publication will get **increasingly obsolete** in semantic publishing environments ...
- ... at least in the STM sector.
- The move into semantic publication will be **much slower in the SSH** because of
  - fuzzy and unstable **terminology**
  - fuzzy **linking semantics** hard to formalise consistently
  - close relation between **complex document formats** and **scholarly discourse**
- Current examples are mostly from the medical and bio-medical area as a consequence
- We are exploring potential of change in the SSH in



## → Much more than catalogues: Visualise and Explore Cultural Context

- Mapping the Republic of Letters:

<http://knot-dev.herokuapp.com/investigate.html>

- Or again the graph of writers and thinkers and how they are connected:

<http://zoom.it/Vj6F> (is this one really useful?)

<http://bgriffen.scripts.mit.edu/www/media/json/thinkers/>

<http://mariandoerk.de/edgemaps/demo/>

<http://www.visualdataweb.org/refinder/refinder.php>

- Or again a Finnish example (Kulttuurisampo):

<http://www.kulttuurisampo.fi/kulsa/historiallisetKartat.shtml>

- Or finally Obama vs. Palin:

[http://truthy.indiana.edu/memedetail?id=324&resmin=45&theme\\_id=4](http://truthy.indiana.edu/memedetail?id=324&resmin=45&theme_id=4) vs.

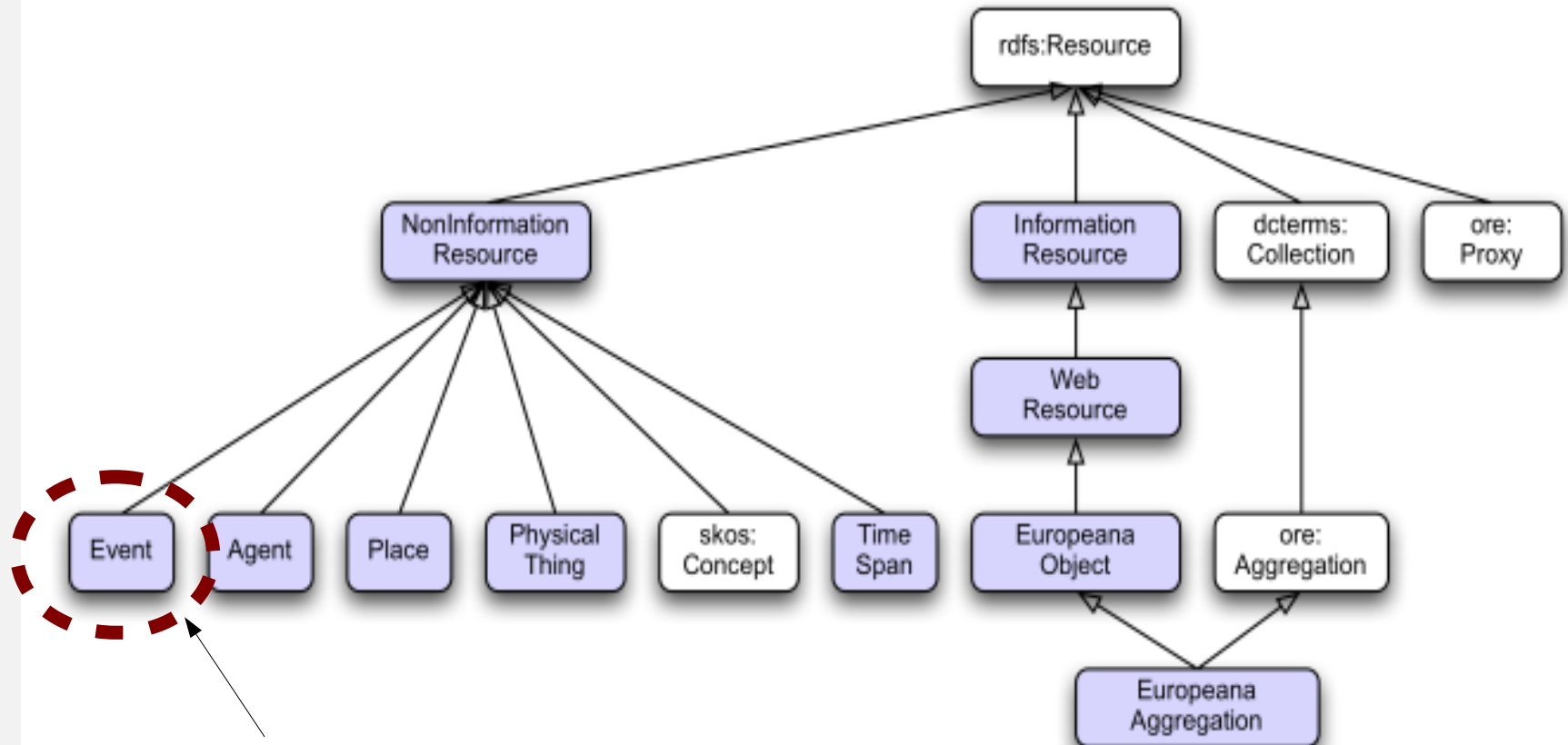
[http://truthy.indiana.edu/memedetail?id=783&resmin=45&theme\\_id=4](http://truthy.indiana.edu/memedetail?id=783&resmin=45&theme_id=4)

# The Europeana Data Model (EDM) in the LoD Context

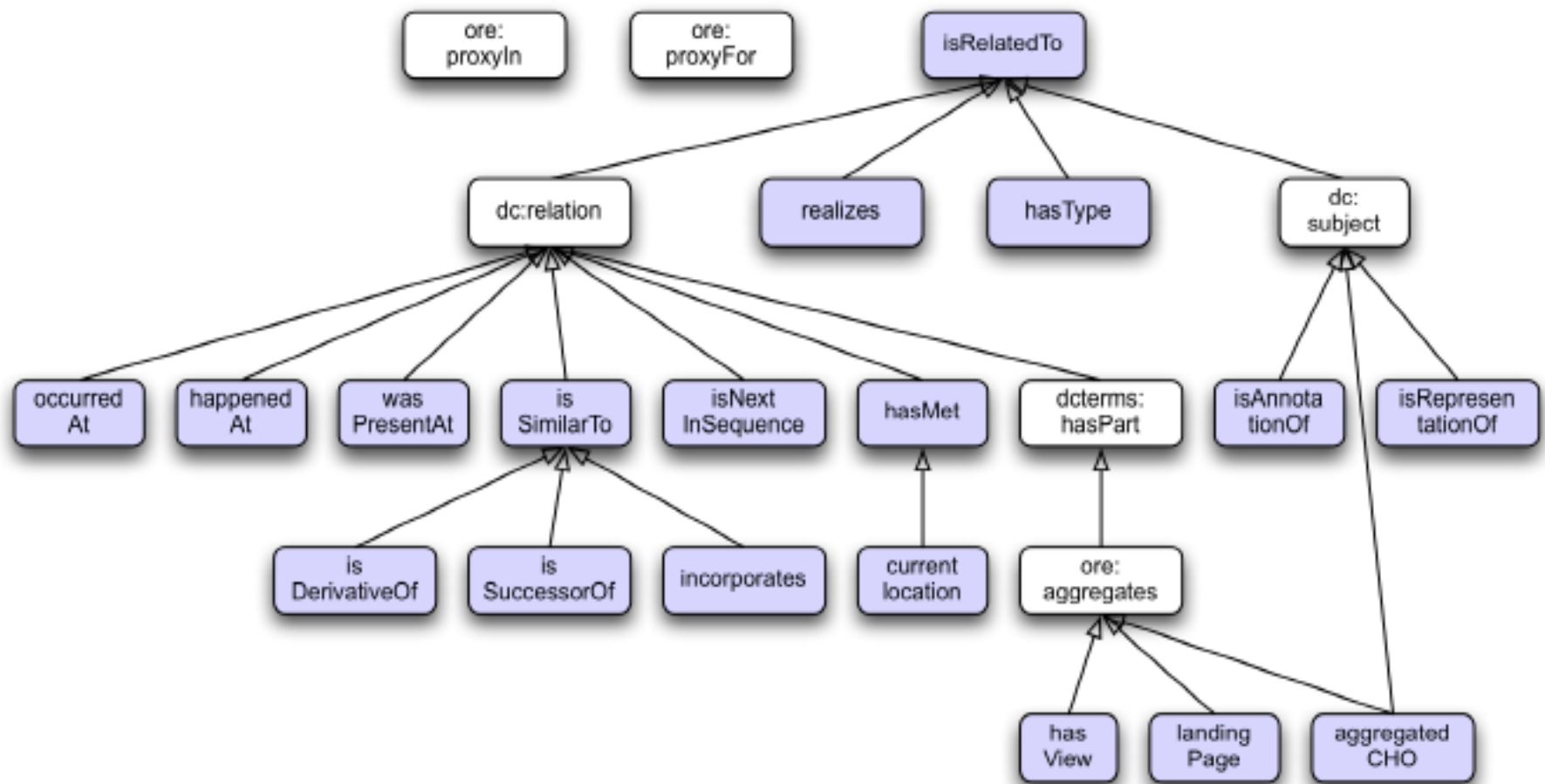
# EDM – what is it?

## And what not?

- EDM is the **metadata model** replacing the ESE ...
- ... a model for making statements about **digital representations** of cultural heritage objects
- ... a model for **contextualising** such representations
- EDM is **not an object model** (but might be combined with object and process models)!
- EDM is an **RDF based graph model**
- *EDM enables modeling of objects and context and thus **knowledge generation***



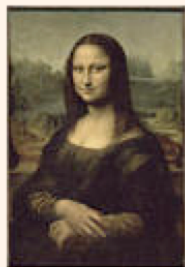
CIDOC CRM E5 hierarchy  
could be pruned here



# Mona Lisa: French Ministry of Culture

KU LEUVEN

Réponse n° 1



Domaine	peinture
Type d'objet	tableau
Titre	<b>PORTRAIT DE MONA LISA (1479-1528) ; DITE LA JOCONDE</b>
Auteur/exécutant	LEONARDO DI SER PIERO DA VINCI ; VINCI Léonard de (dit)
Précision auteur/exécutant	Vinci, 1452 ; Amboise, 1519
Ecole	Italie
Période création/exécution	1er quart 16e siècle
Millésime création/exécution	1503 entre ; 1506 et
Genèse	oeuvre en rapport ; reproduit en gravure
Historique	commandé par le florentin Francesco del Giocondo, époux de Mona Lisa entre 1503 et 1506 ; nombreuses copies dont une conservée au Louvre ; gravé par Fauchery, par Filhol, par Landon
Matériaux/techniques	peinture à l'huile ; bois
Mesures	77 H ; 53 L
Sujet représenté	portrait (Mona Lisa, femme, à mi-corps, de trois-quarts, assis, accoudé, loggia, Italien) ; fond de paysage (montagne, rocher, cours d'eau, pont, plaine, route)
Date sujet représenté	1479-1528
Lieu de conservation	Paris ; musée du Louvre département des Peintures
	 <i>Musée de France</i> <i>au sens de la loi n° 2002-5 du 4 janvier 2002</i>
Statut juridique	propriété de l'Etat ; musée du Louvre département des Peintures
Anciennes appartenances	François 1er ; Couronne de France
Numéro d'inventaire	INV 779
Commentaires	légère diminution du tableau sur les côtés (environ 7 mm) ; acheté vraisemblablement vers 1519, après la mort de l'artiste
Bibliographie	HEYDENRICH 6 ; OTTINO DELLA CHIESA 31 ; VILLOT I 484 ; HAUTECOEUR 1601 ; C.S.I. 1981, P 192
Copyright notice	© Musée du Louvre, © Direction des Musées de France, 1999
Crédits photographiques	© Réunion des musées nationaux ; © Hervé Lewandowski ; © Thierry Le Mage
	 commande reproduction et/ou conditions d'utilisation
	renseignements sur le musée
	000PE025604

**KU LEUVEN**



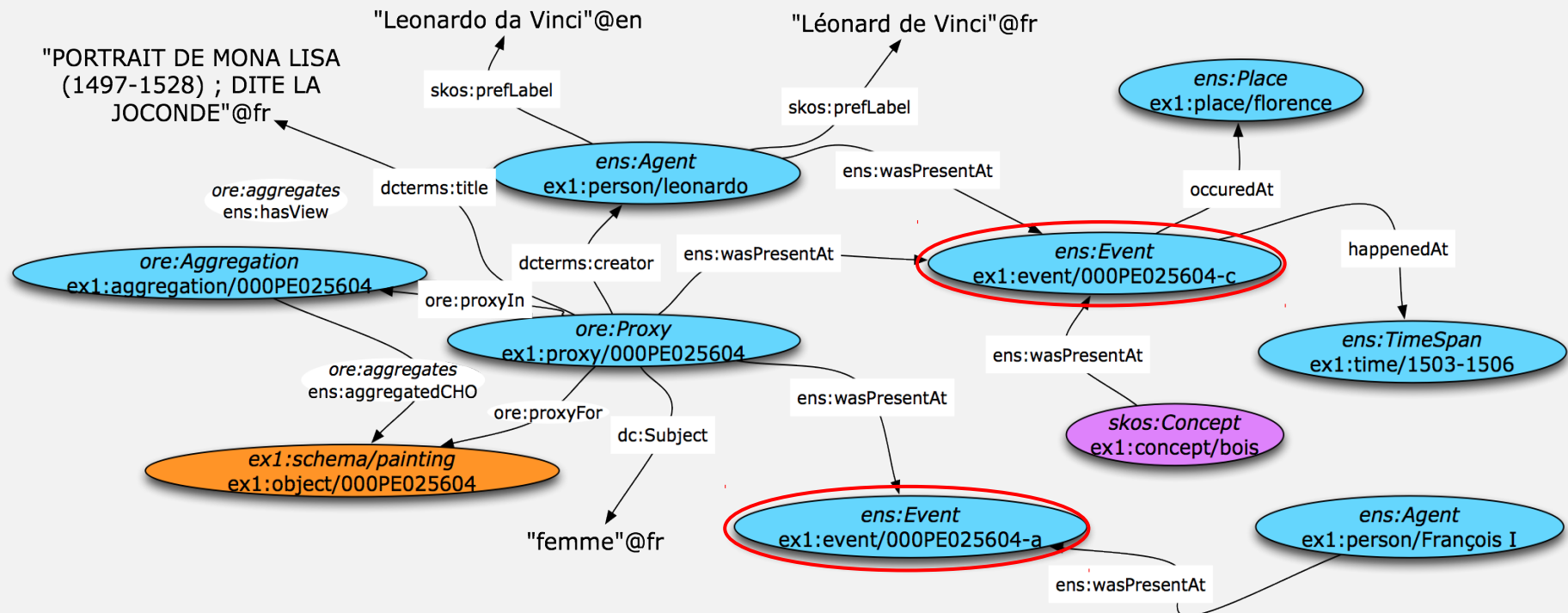


**skos:Concept:** entities from KOS

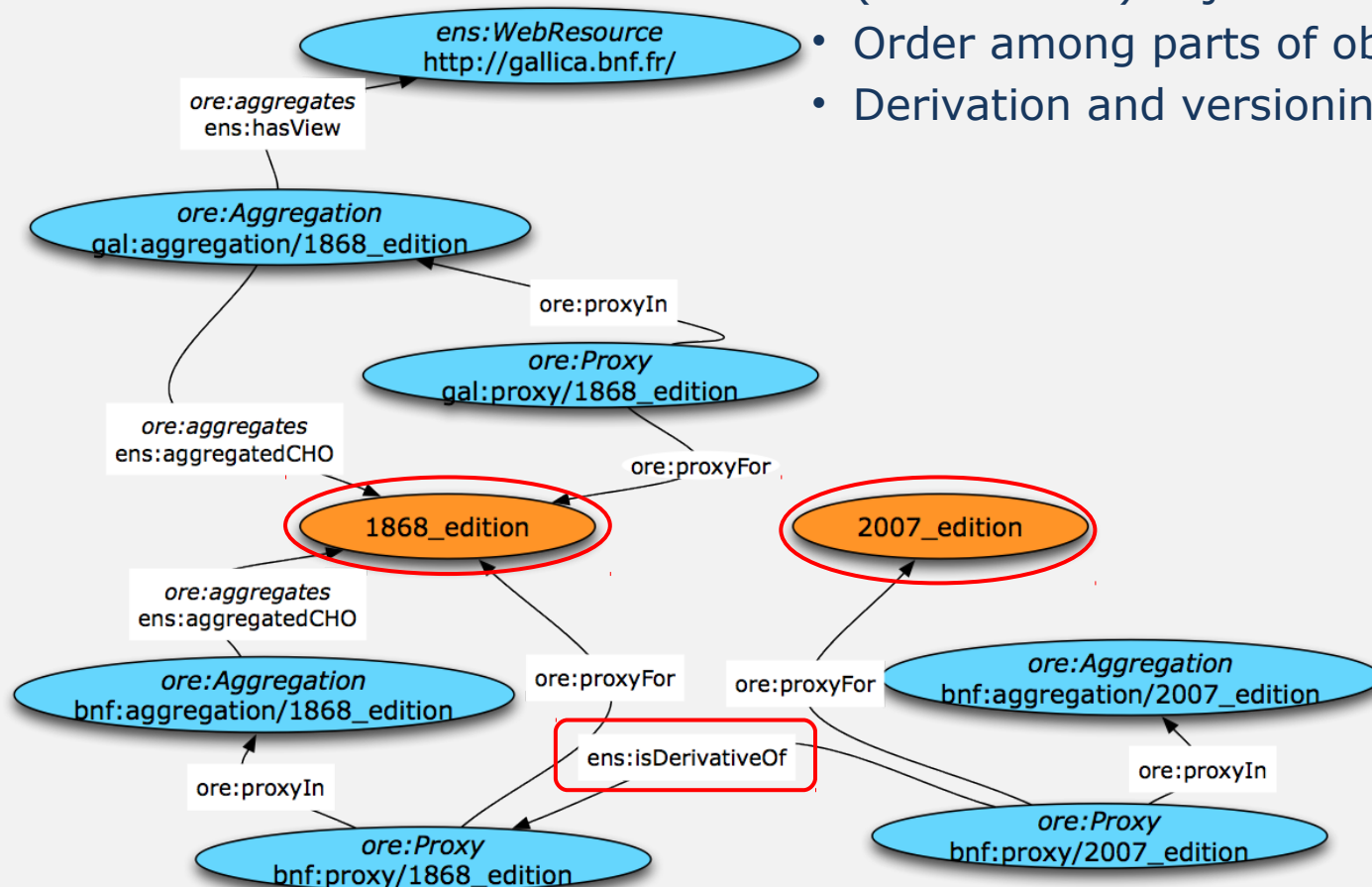
"Léonard de Vinci"@fr



Preserving and exploiting original data also means being compatible with descriptions beyond simple object level (→ CIDOC CRM!)



- Part-whole links for complex (hierarchical) objects
- Order among parts of objects
- Derivation and versioning relations



# Les Fleurs du Mal: UNIMARC

KU LEUVEN

<http://catalogue.bnf.fr/ark:/12148/cb37367035f>

```
000 nam 22 450
001FRBNF373670350000003
009http://catalogue.bnf.fr/ark:/12148/cb37367035f
039 $oGEA$a000288182
100 $a19920409d1857 m y0frey50 ba
1010 $afre
102 $aFR
105 $a||||z 00|||
106 $ar
2001 $a^Les %fleurs du mal$bTexte imprimé$fpar Charles
Baudelaire
210 $aParis$cPoulet-Malassis et De Broise$d1857
215 $a248 p.$d19 cm
676 $a841.8$v22
686 $a840$2Cadre de classement de la Bibliographie nationale
française
700 |$311890582$aBaudelaire$bCharles$4070
801 0$aFR$bBNF$c19920409$gAFNOR$2intermrc
```



# Les Fleurs du Mal: Gallica

<http://gallica.bnf.fr/ark:/12148/bpt6k70861t>

KU LEUVEN

Gallica Consultation : Les fleurs du mal / par... page Non paginé of 258 (Display: Normal)

http://gallica.bnf.fr/ark:/12148/bpt6k70861t#

BnF gallica Digital library

Home > Consultation

Les fleurs du mal / par Charles Baudelaire  
Baudelaire, Charles, 1857  
Full record  
Display plain text

Search this document

Browse by

Page  
Thumbnail  
Table of contents

NP (not numbered)  
NP (not numbered)  
NP (not numbered)  
NP (not numbered)  
NP (not numbered)  
p.1  
p.2  
p.3  
p.4  
p.5  
p.6  
p.7  
p.8

Full record

Title : Les fleurs du mal / par Charles Baudelaire  
Author : Baudelaire, Charles  
Publisher : Poulet-Malassis et De Broise (Paris)  
Date of publication : 1857  
Type : monographie imprimée  
Language : French  
Format : 248 p. ; 19 cm  
Format : application/pdf  
Copyright : domaine public  
Identifier : ark:/12148/bpt6k70861t  
Source : Bibliothèque nationale de France, Rés. Ye-3535  
Relation : <http://catalogue.bnf.fr/ark:/12148/cb37367035f/description>  
Provenance : bnf.fr  
Theme : French & related literatures

Close

Search

OK

Advanced search

Custom use

Assemble your own digital library and tag your documents

Sign in  
create an account

About this document

Add a Tag  
Add this document

Permalink on this document  
<http://gallica.bnf.fr/ark:/12148/bpt6k70861t>

Entfernen

Fertig

Downloads HybridPubl\_in...

# Les Fleurs du Mal: Digitised

KU LEUVEN

<http://gallica.bnf.fr/ark:/12148/bpt6k70861t.textePage.f1>

Text mode | Audio mode

page NP (screen 1 of 258)

[Next page](#) | [Last page](#) |

**Title :** Les fleurs du mal / par Charles Baudelaire

**Author :** Baudelaire, Charles

**Publisher :** Poulet-Malassis et De Broise (Paris)

**Date of publication :** 1857

**Type :** monographie imprimée

**Language :** French

**Format :** 248 p. ; 19 cm

**Format :** application/pdf

**Copyright :** domaine public

**Identifier :** [ark:/12148/bpt6k70861t](http://gallica.bnf.fr/ark:/12148/bpt6k70861t)

**Source :** Bibliothèque nationale de France, Rés. Ye-3535

**Relation :** <http://catalogue.bnf.fr/ark:/12148/cb37367035f/description>

**Provenance :** bnf.fr

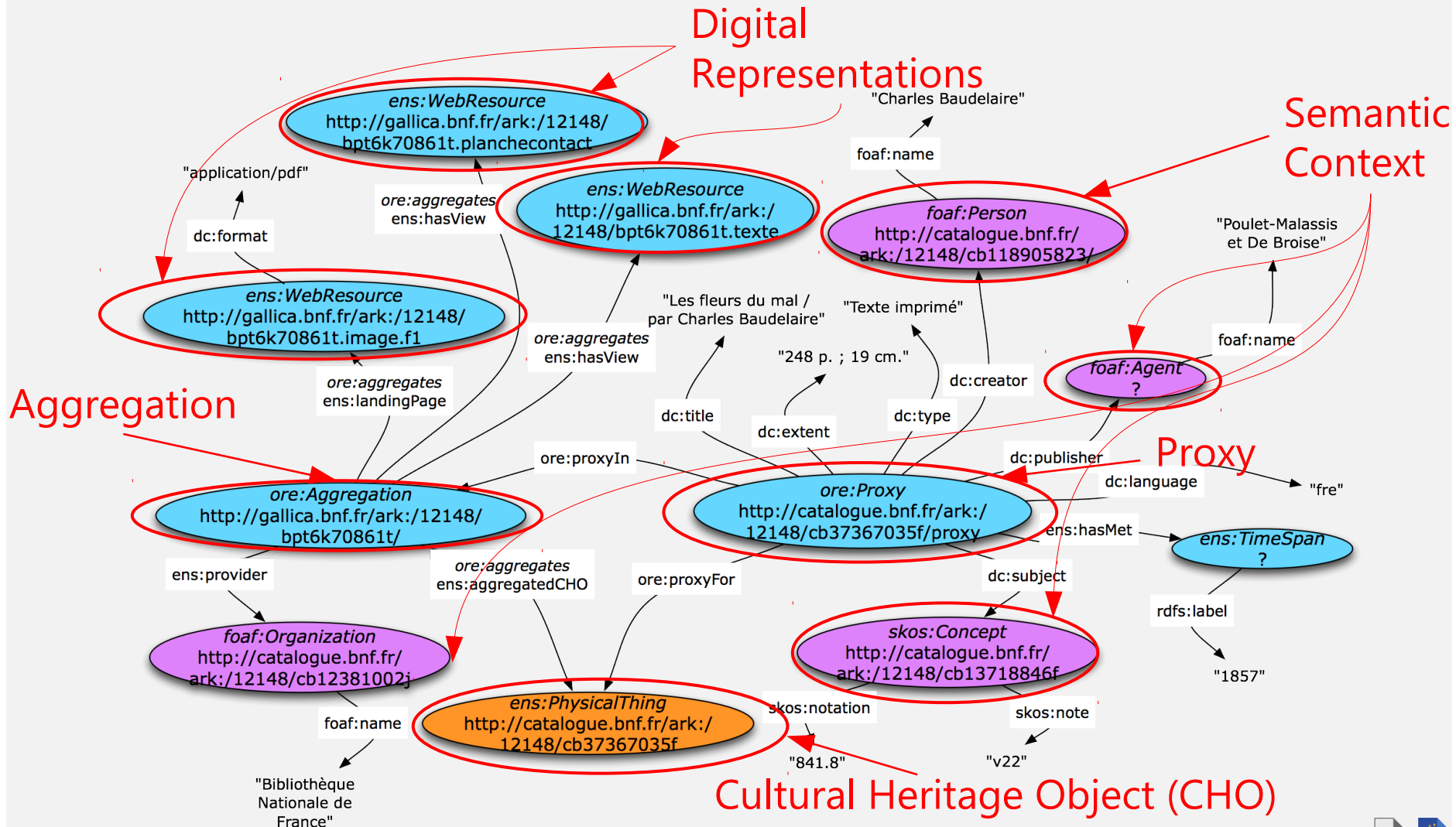
**Theme :** French & related literatures

*The text below has been produced using a process called optical character recognition (O.C.R.). Since it is an automatic process, it is subject to errors you might find in this page.*

*The recognition rate for this document is 94.63 %.*



# Les Fleurs du Mal: EDM





# What can you use it for:

## De arte venandi cum avibus

# De Arte Venandi ... (1)

KU LEUVEN

## Frederick II, Holy Roman Emperor: De arte venandi cum avibus. Moamin. Gerardus Falconarius. Ghatrif. Grisofus medicus



Shelfmark	Valencia Universitat de València, Biblioteca Històrica BH Ms. 601
Author	Frederick II, Holy Roman Emperor, 1194-1250 Gerardus, Falconarius Ghatrif Grisofus, Medicus
Date of Origin	around 1470
Place of Origin	Italy
Material	parchment
Extent	304 f.
Dimensions	329 x 211 mm
Illuminated	yes
Language	Latin

# De Arte Venandi ... (2)

KU LEUVEN

Bildschirmfoto Ablage Bearbeiten Foto Fenster Hilfe

Europeana Profes: x VIVO | connect - x edoc - Dokument x COST/SISMEL Trai x Europeana Regia x Google Kalender x Protokoll Besprecl x crossref.org : : pu x

www.europeana-regia.eu/en/manuscripts/authors/frederick-ii-holy-roman-emperor-1194-1250

Erste Schritte Apple Neuer Ordner DM2E Pundit Import to Mendeley Yahoo! Google Maps YouTube Wikipedia News Beliebt » Andere Lesezeichen

CA DE EN ES FR NL

europaana Regia

HOME PROJECT EUROPEANA REGIA HISTORICAL COLLECTIONS MANUSCRIPTS NEWS LINKS

## Frederick II, Holy Roman Emperor, 1194-1250

View International Authority File ID Number: 66596013  
URL: <http://viaf.org/viaf/66596013/>

FOLLOW US

Manuscripts

News

Facebook

Twitter

HOME | TERMS OF USE | CONTACT

25

# De Arte Venandi ... (3)

KU LEUVEN

Bildschirmfoto Ablage Bearbeiten Foto Fenster Hilfe

Europeana Prof x VIVO | connect x Edoc - Dokume x COST/SISME T x Europeana Reg x VIAF x Google Kalendi x Protokoll Bespr x crossref.org : x

viaf.org/viaf/66596013/

Erste Schritte Apple Neuer Ordner DM2E Pundit Import to Mendeley Yahoo! Google Maps YouTube Wikipedia News Beliebt » Andere Lesezeichen

## VIAF




Virtual International Authority File


### Suchen


Feld auswählen: Index auswählen: Suchbegriffe:


Gesamter VIAF

Suchen

**Frederick II, Holy Roman Emperor, 1194-1250**   

**Frédéric II empereur germanique 1194-1250** 















**1250-1194 فردريك الثاني، الإمبراطور الروماني المقدس،** 

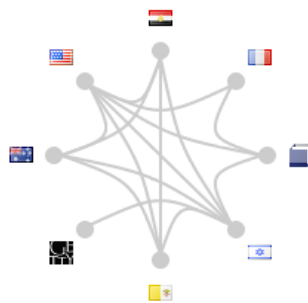
**Frederick II, Holy Roman Emperor Roman emperor, 1194-1250** 

VIAF ID: 66596013 (Personal)

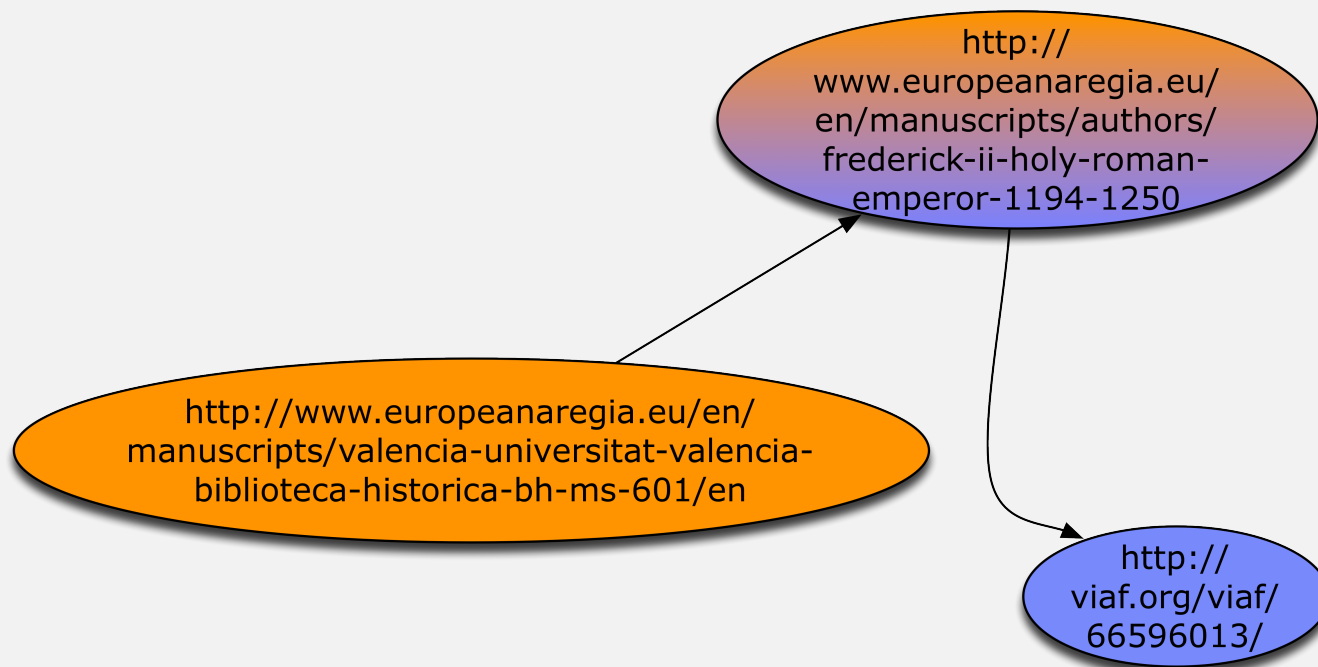
Permalink: <http://viaf.org/viaf/66596013>

### Vorzugsbezeichnungen

-   100 0 \_ [\\_ta Frederick \\_tb II, \\_tc Holy Roman Emperor, \\_td 1194-1250](#)
-   100 0 \_ [\\_ta Frédéric \\_tb II, \\_tc empereur germanique, \\_td 1194-1250](#)
-   200 \_ | [\\_ta Frédéric \\_td II \\_tc empereur germanique \\_tf 1194-1250](#)
-   100 0 \_ [\\_ta فردريك \\_tb الثاني, \\_tc الإمبراطور الروماني المقدس, \\_td 1194-1250](#)
-   100 0 \_ [\\_ta Frederick \\_tb II, \\_tc Holy Roman Emperor, \\_td 1194-1250](#)
-   100 0 \_ [\\_ta Frederick \\_tb II, \\_tc Holy Roman Emperor, \\_td 1194-1250](#)
-   100 1 \_ [\\_ta Frederick II, Holy Roman Emperor \\_tg Roman emperor, 1194-1250](#)



# De Arte Venandi ... Subgraph 1





# De Arte Venandi ... (4)

KU LEUVEN

Bildschirmfoto Ablage Bearbeiten Foto Fenster Hilfe 96 % Di. 10:30 gradmans

Europeana Profes: x VIVO | connect - x Edoc - Dokument x COST/SISEMEL Trai x Europeana Regia x Google Kalender x Protokoll Besprecl x crossref.org : : pu x

www.europeana-regia.eu/en/manuscripts/places/italy

Erste Schritte Apple Neuer Ordner DMZE Pundit Import to Mendeley Yahoo! Google Maps YouTube Wikipedia News Beliebt » Andere Lesezeichen

CA DE EN ES FR NL

HOME PROJECT EUROPEANA REGIA HISTORICAL COLLECTIONS MANUSCRIPTS NEWS LINKS

## Italy

Getty Thesaurus of Geographic Names ID Number: 1000080  
URL: [http://www.getty.edu/vow/TGNFullDisplay?find=1000080&place=&nation=&prev\\_page=1&...](http://www.getty.edu/vow/TGNFullDisplay?find=1000080&place=&nation=&prev_page=1&...)

FOLLOW US

Manuscripts

News

Facebook

Twitter

HOME | TERMS OF USE | CONTACT

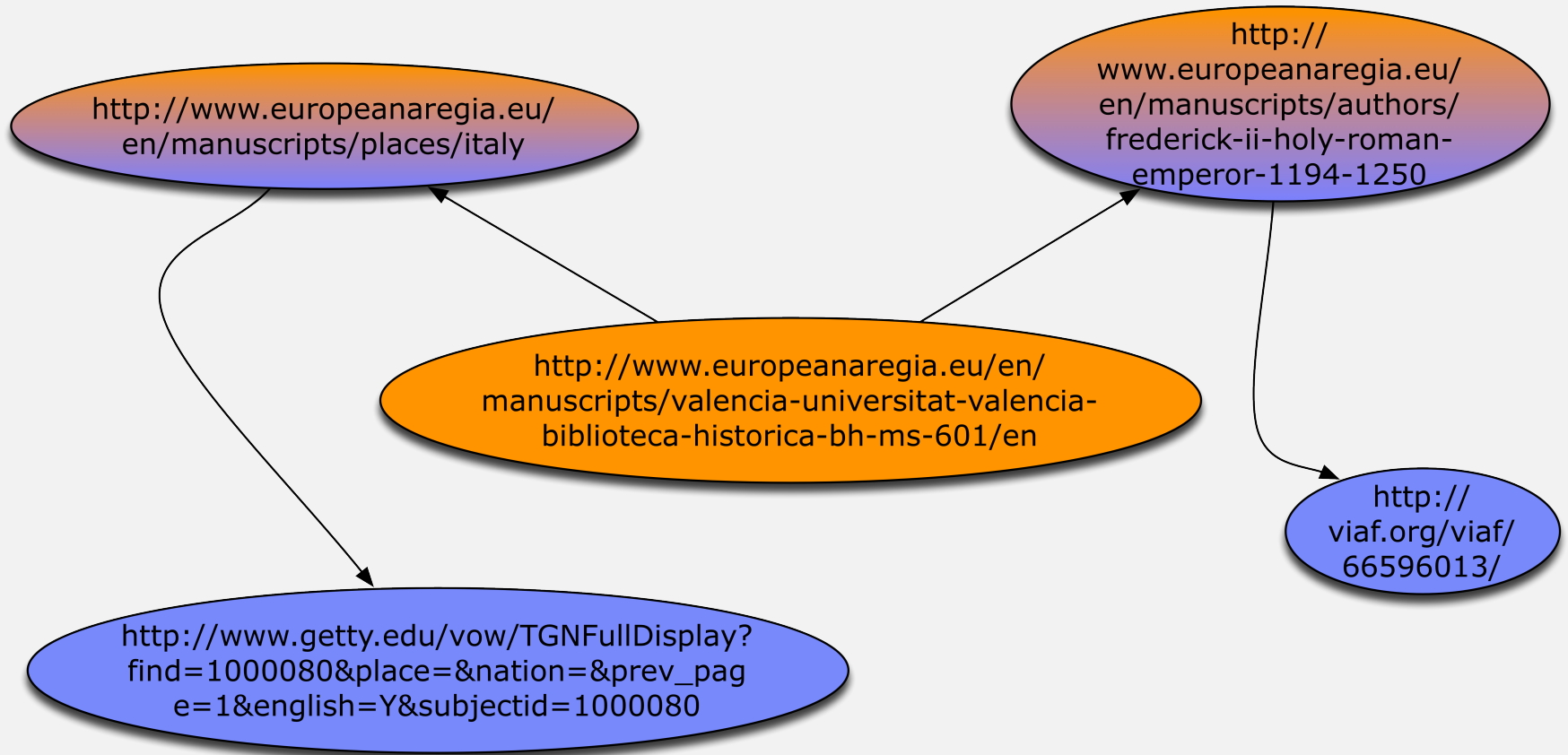
Mac OS X desktop icons: Finder, Launchpad, Safari, Mail, iCal, iPhoto, iTunes, etc.

**KU LEUVEN**





# De Arte Venandi ... Subgraphs 1+2



## Frederick II, Holy Roman Emperor: De arte venandi cum avibus. Moamin. Gerardus Falconarius. Ghatrif. Grisofus medicus



Shelfmark	Valencia Universitat de València, Biblioteca Històrica BH Ms. 601
Author	Frederick II, Holy Roman Emperor, 1194-1250 Gerardus, Falconarius Ghatrif Grisofus, Medicus
Date of Origin	around 1470
Place of Origin	Italy
Material	parchment
Extent	304 f.
Dimensions	329 x 211 mm
Illuminated	yes
Language	Latin

# De Arte Venandi ... (6)

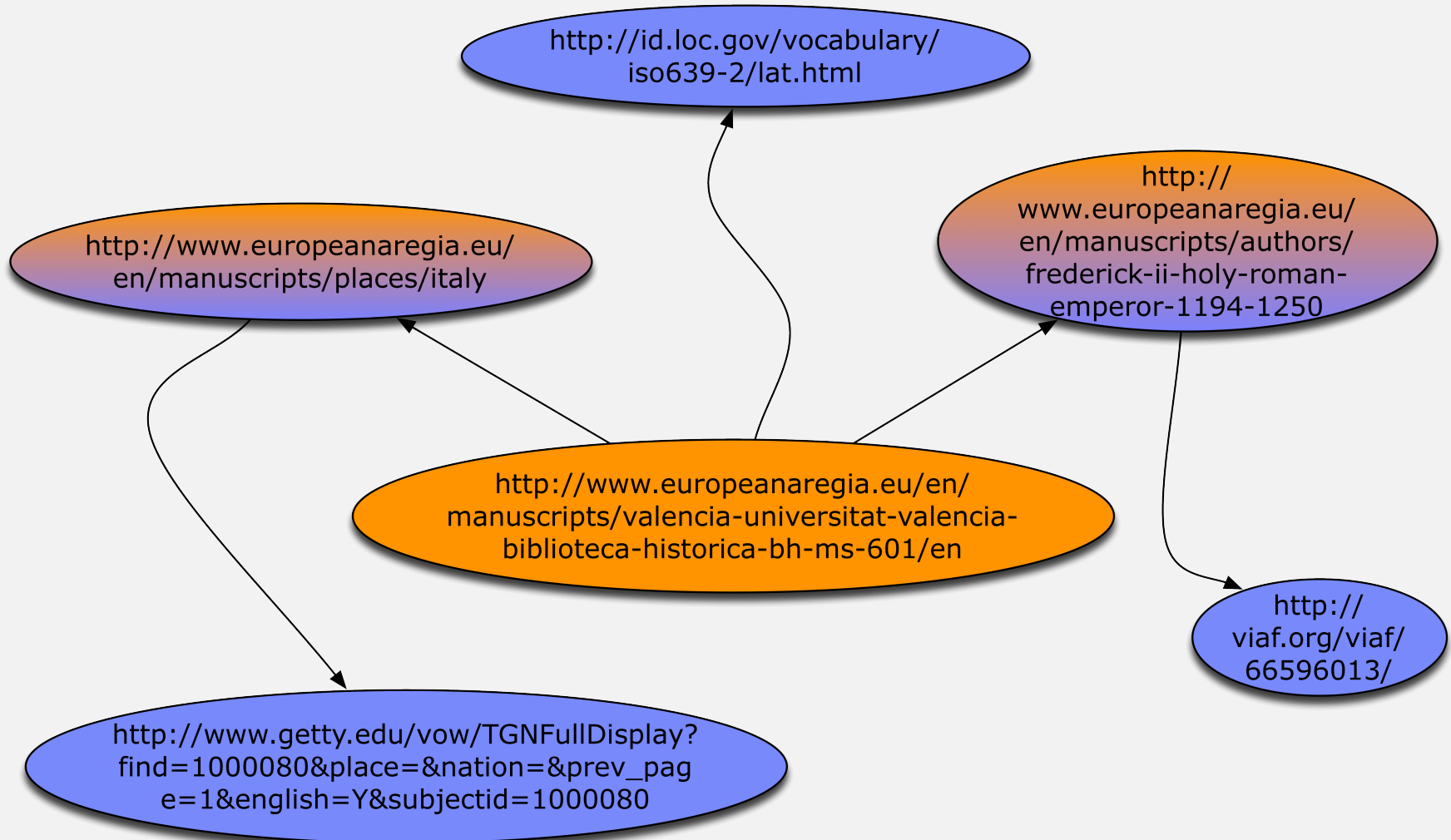
KU LEUVEN

The screenshot shows a web browser window with the address bar displaying `id.loc.gov/vocabulary/iso639-2/lat.html`. The browser's menu bar includes 'Bildschirmfoto', 'Ablage', 'Bearbeiten', 'Foto', 'Fenster', and 'Hilfe'. The toolbar shows various icons for file operations and search. The browser's tab bar contains several open tabs, including 'Europeana Prof', 'VIVO | connect', 'edoc - Dokume', 'COST/SISMEL T', 'Europeana Reg', 'Latin; latin - LC', 'Google Kalend', 'Protokoll Bespr', and 'crossref.org'. The browser's status bar at the bottom shows the system clock as 'Di. 10:37' and the user as 'gradmans'. The main content area of the browser displays the Library of Congress website. The header of the website includes the 'LIBRARY OF CONGRESS' logo, navigation links for 'ASK A LIBRARIAN', 'DIGITAL COLLECTIONS', and 'LIBRARY CATALOGS', and a search bar with the text 'Search Loc.gov' and a 'GO' button. Below the header, a breadcrumb trail reads 'The Library of Congress > Linked Data Service > ISO639-2 Languages'. The main heading is 'Latin; latin'. Below this, a link points to 'From ISO 639-2: Codes for the Representation of Names of Languages - Part 2: Alpha-3 Code for the Names of Languages'. The page features two tabs: 'Details' (selected) and 'Visualization'. Under the 'Details' tab, the following information is displayed: 'Latin' with a flag icon, 'latin' with a flag icon, 'Bibliographic Code', 'URI(s)' with a link to `http://id.loc.gov/vocabulary/iso639-2/lat`, 'Instance Of' with links to 'MADS/RDF Language', 'MADS/RDF Authority', 'SKOS Concept', 'RDF Resource', and 'ISO6392 Language', 'Scheme Membership(s)' with a link to 'ISO 639-2: Codes for the Representation of Names of Languages - Part 2: Alpha-3 Code for the Names of Languages', and 'Collection Membership(s)' with links to 'http://id.loc.gov/vocabulary/iso639-2/collection\_iso639-2\_Bibliographic\_Codes', 'http://id.loc.gov/vocabulary/iso639-2/collection\_PastPresentISO639-2Entries', and 'http://id.loc.gov/vocabulary/iso639-2/collection\_iso639-2'.

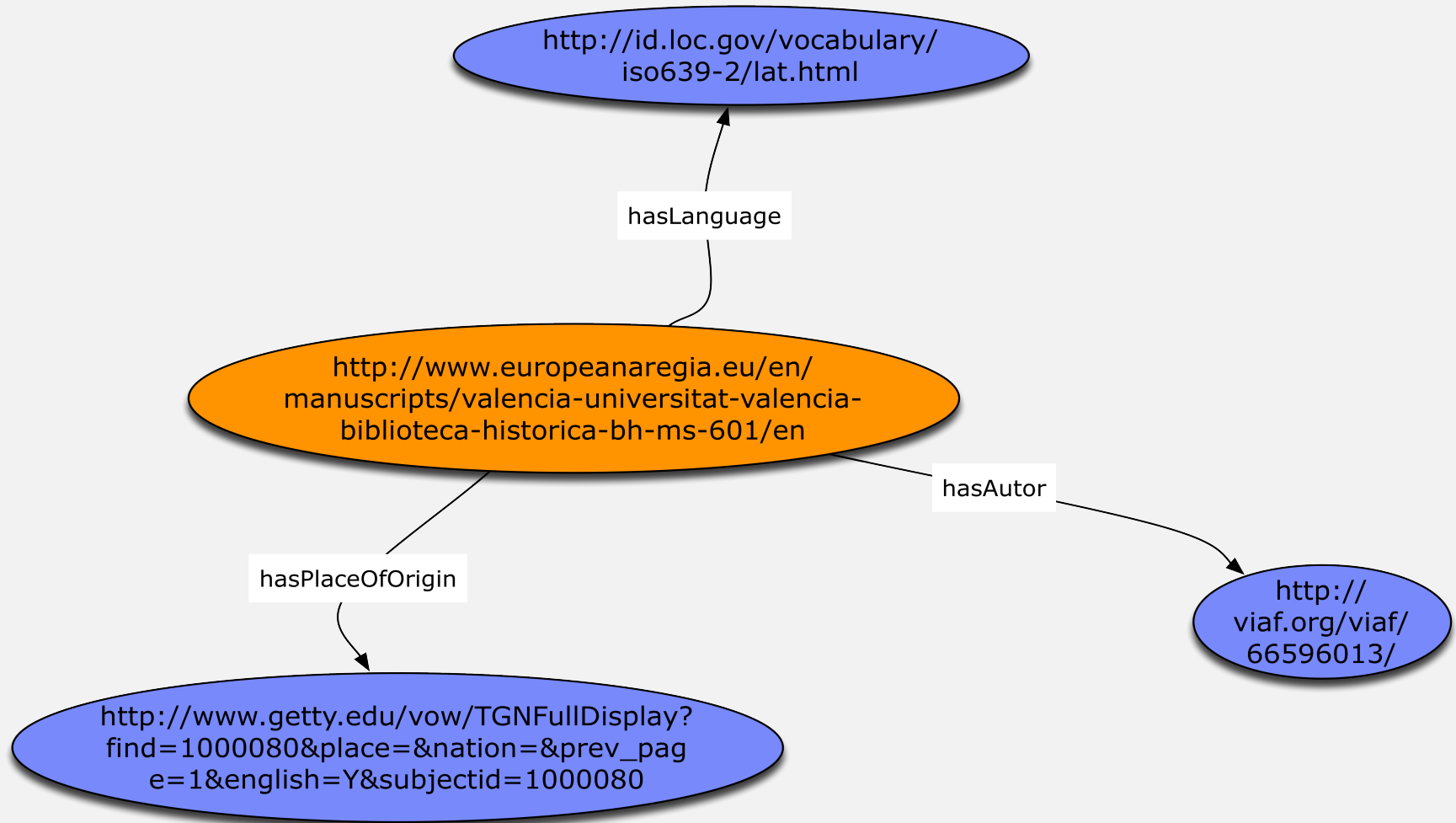
From Records to Graphs: Linked Data and Libraries

Stefan Gradmann, Universidad Carlos III de Madrid, 11/07/2013

# De Arte Venandi ... (6)

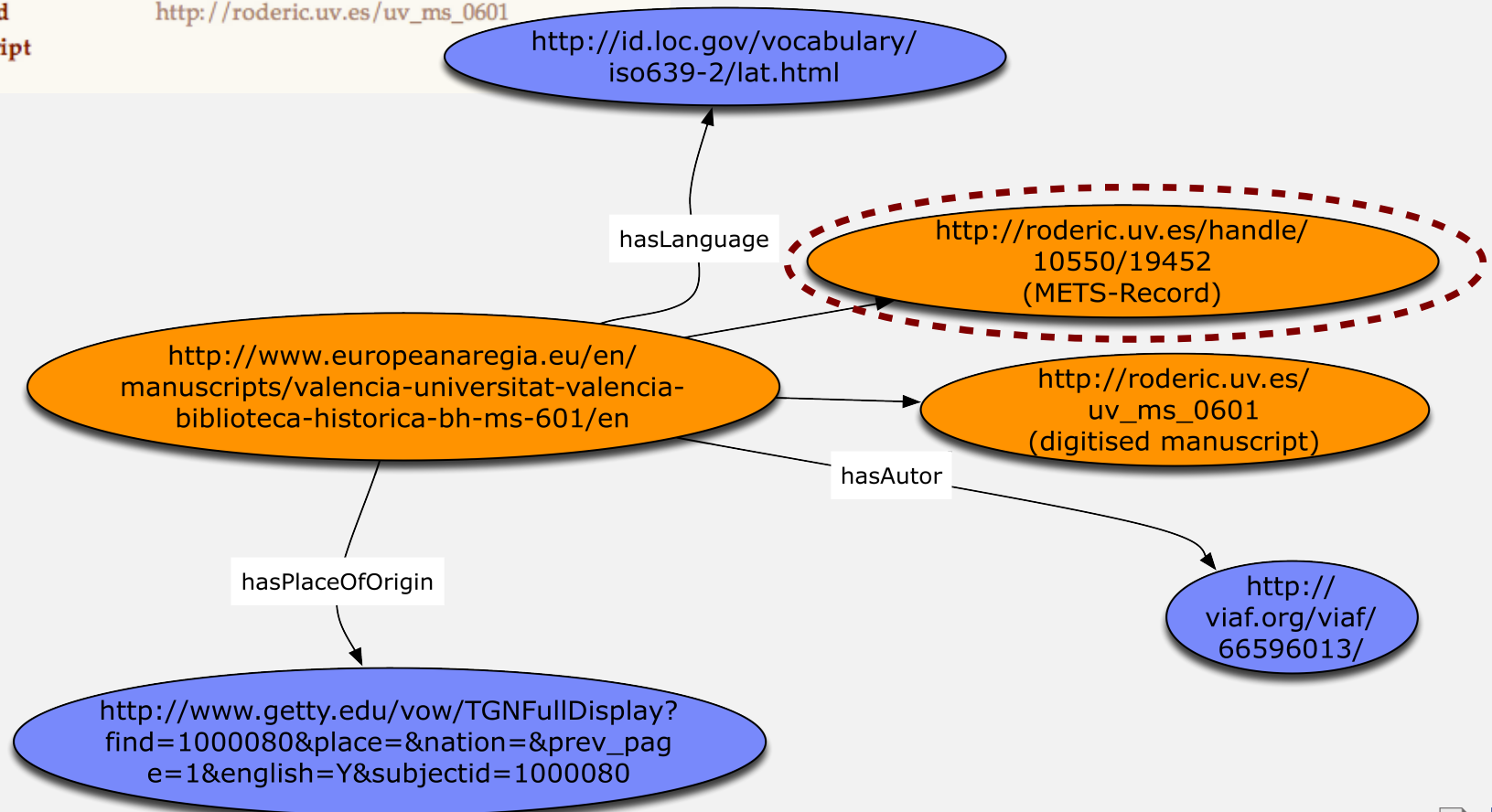


# De Arte Venandi ... done 'right'



# De Arte Venandi ... there's more!

Full Manuscript Record  
<http://trobes.uv.es/record=b1948745>  
<http://roderic.uv.es/handle/10550/19452>  
Digitised Manuscript  
[http://roderic.uv.es/uv\\_ms\\_0601](http://roderic.uv.es/uv_ms_0601)





# De Arte Venandi ...

there

<http://id.loc.gov/vocabulary/iso639-2/lat.html>

xmlui.dri2xhtml.METS-1.0.item-somn[Prologus] "[text] Presens opus agendi nos induxit instans tua petitio  
vir clarissime, M. S. ... nobiles delectantur, vita comite, post  
complementu[m] huius operis diceatur a nobis. [rúbr.] Divi augusti  
Federici secundi romanor  
regis super librum De avibus [et] de occupatione prola  
2. f. 5r-56r: "[rúbr.] Incipit[ue] eius tractat[us] [et] primo de divisione  
g[e]n[er]aliter aviu[m] in aquaticis terrestribus [et] mediis. Item in  
rapacibus [et] non rapacibus. [text] In hoc autem primo tractatu huius  
operis dicemus generaliter de avibus ...; [f. 55r] [et] que dicuntur  
anates campestres [et] plures alie ... [f. 56r] [rúbr.] Expicitus est  
primus tractatus"

ore:Aggregation  
De Arte Venandi ...

3. f. 56r-115r: "[rúbr.] Incipit[ur] de divisione  
particulis. ... [et] de posterior  
venatio... Ex  
est q[uam] paulatina  
4. f. 115r-145r: "[rúbr.] Incipit[ur] [ter]cius. Dictum est  
mansuefactione falconu[m] cum capello. Seq[ui]tur  
instrumentis p[er] quos redeat ad homines. [text] Intentio n[on]  
principio diximus est docere venationes ... [et] ledere etiam ipsos in  
p[er]sona. [rúbr.] Explicit liber quartus [tat]at[ur] [ter]cius"  
5. f. 145r-205v: "[rúbr.] Incipit[ur] [qu]artus. De girofalcone ad gruem,  
que est quondam in venatione quondam eo. [text]  
ad grues ... qua[m] habent modi falcomu[m] [rúbr.]  
Explicit liber quartus de girofalcis ad grues"  
6. f. 174v-205v: "[rúbr.] Incipit liber quintus. De falcone sacro ad  
ayrones. [text] Determinatum est in precedentib[us] sufficienter  
girofalcis ... Et grus volans percutit falconem ayro aute  
Dictum est de girofalcone ad grues. Explic

[http://roderic.uv.es/uv\\_ms\\_0601](http://roderic.uv.es/uv_ms_0601)  
(digitised manuscript)

[http://roderic.uv.es/handle/uv\\_ms\\_0601/1-f-3r-5r.html](http://roderic.uv.es/handle/uv_ms_0601/1-f-3r-5r.html)  
(METS-Record)

edm:isNextInSequence

[http://roderic.uv.es/handle/uv\\_ms\\_0601/2-f-5r-56r.html](http://roderic.uv.es/handle/uv_ms_0601/2-f-5r-56r.html)  
(METS-Record)

edm:isNextInSequence

[http://roderic.uv.es/handle/uv\\_ms\\_0601/3-f-56r-115r.html](http://roderic.uv.es/handle/uv_ms_0601/3-f-56r-115r.html)  
(METS-Record)

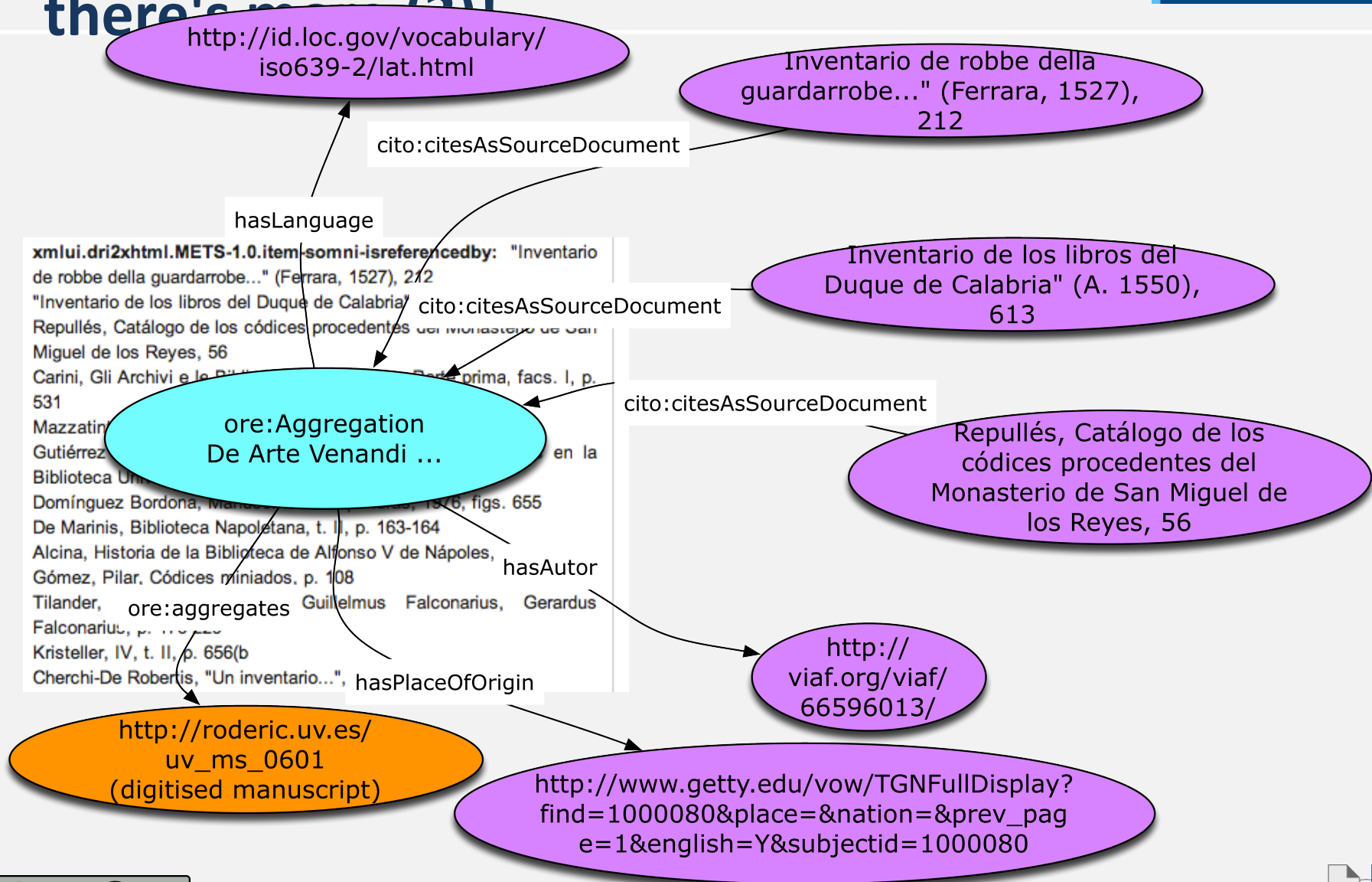
<http://viaf.org/viaf/66596013/>

[http://www.getty.edu/vow/TGNFullDisplay?find=1000080&place=&nation=&prev\\_page=1&english=Y&subjectid=1000080](http://www.getty.edu/vow/TGNFullDisplay?find=1000080&place=&nation=&prev_page=1&english=Y&subjectid=1000080)

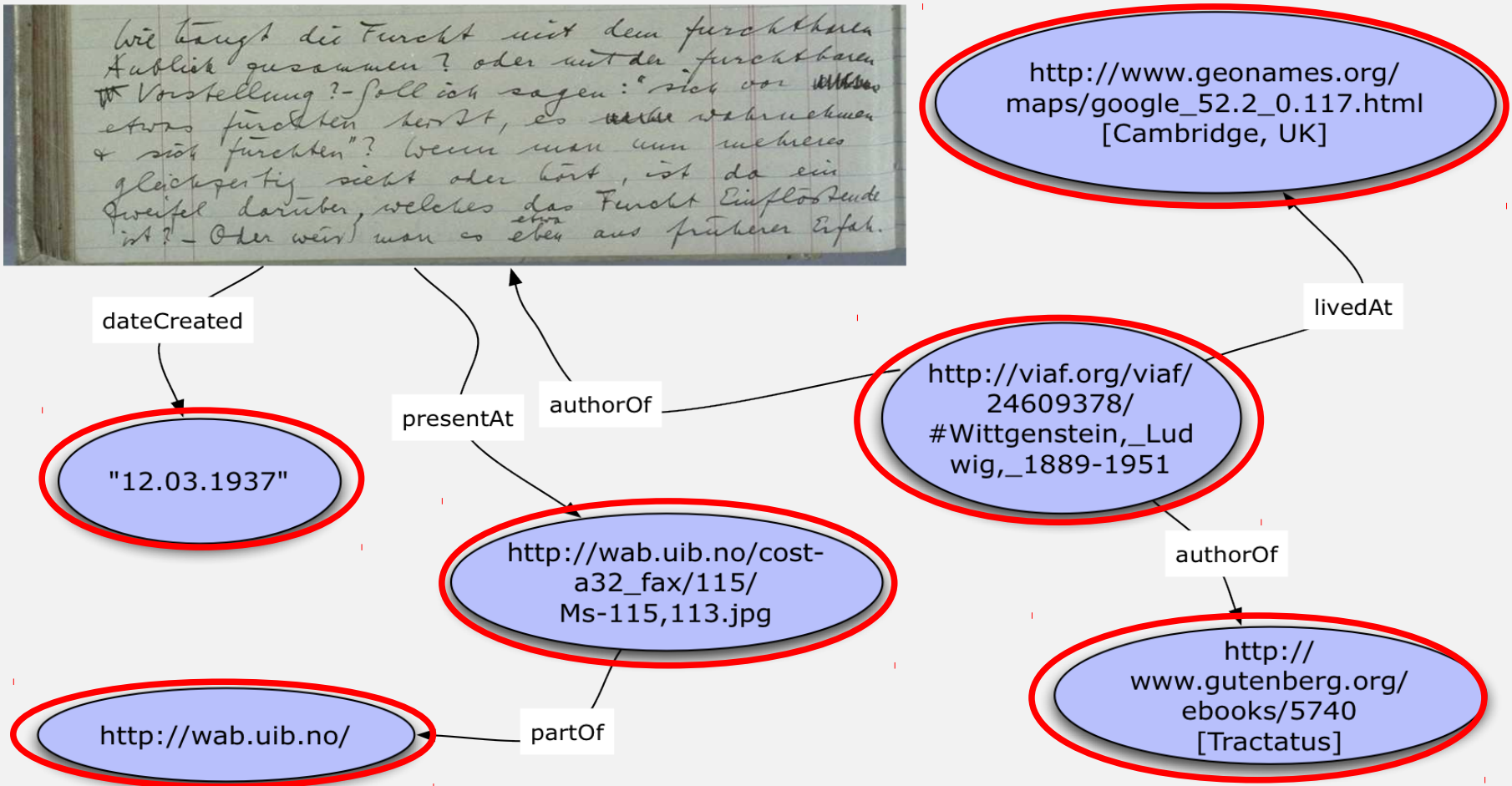
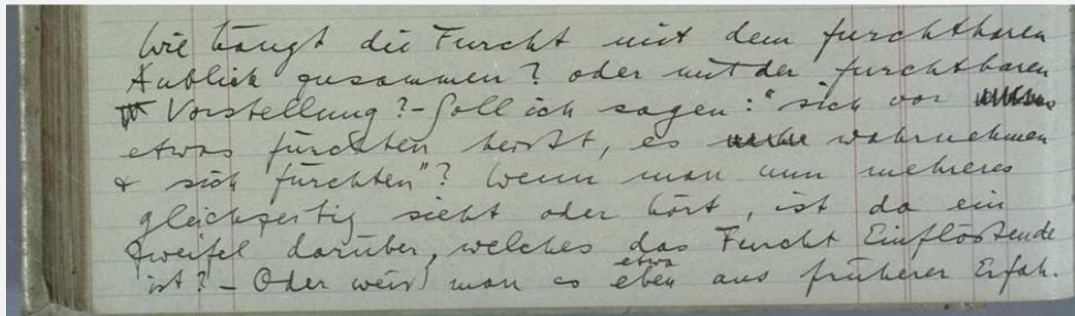


# De Arte Venandi ...

## there's more (2)



# Contextualising Wittgenstein



# An Opportunity for the Library ...

## ... and what it needs to do to be up to it

# “What do you do with a million books?” (Greg Crane)

- Digitisation and semantic publishing result in
  - growing quantity
  - increased complexity
- Well beyond scholarly processing capacity (=reading faculty)
- Scientists and Scholars will badly need help in three areas:
  - **Semantic abstracting, named entity recognition** for “strategic reading” (Renear)
  - **Contextualisation** of information objects
  - Robust **reasoning and inferencing** yielding digital heuristics
- => **Opportunities for Research Libraries!**

# Ceci n'est pas une bibliothèque





# Ceci n'est pas une bibliothèque



# Catalogue



The card catalog in the nave of Sterling Memorial Library at Yale University. Picture by Henry Trotter, 2005.



# Catalogue Entry: MARC Record

**LEADER:** 00495cam 2200169 a 4500  
001 94205672  
003 DLC  
005 2004010812243.0  
008 940907s1926 gw 000 1 ger  
010 \$a 94205672  
040 \$aDLC\$cDLC\$dDLC  
050 00 \$aPT2621.A26\$bS36 1926  
100 1 \$aKafka, Franz,\$d1883-1924  
245 14 \$aDas Schloss :\$bRoman /\$cFranz Kafka.  
260 \$aMünchen :\$bKurt wolff verlag,\$c1926.  
300 \$a505 p. ;\$c20 cm.  
561 \$aSource: Bequest of Aaron Copland, Fall, 1989.\$5DLC

# 'Library Collections'

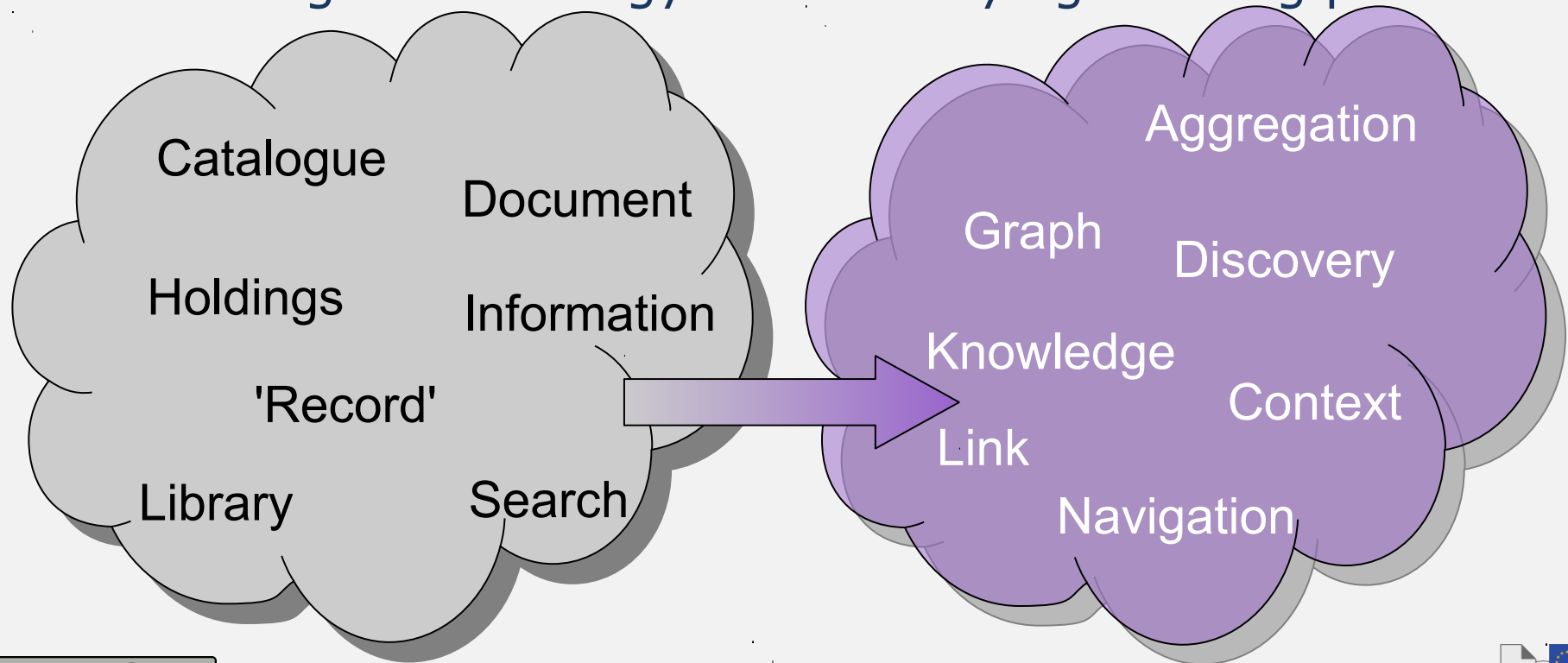


Photo © Ralf Küpper



# Change Thinking, Change Terminology!

- Libraries will serve research as part of the Linked Open Data web – or else risk becoming insignificant.
- For operating this change we definitely need to change terminology and underlying thinking patterns:



# Sticking to empty metaphors ...

"What's in a name? That which we call a rose  
By any other name would smell as sweet."

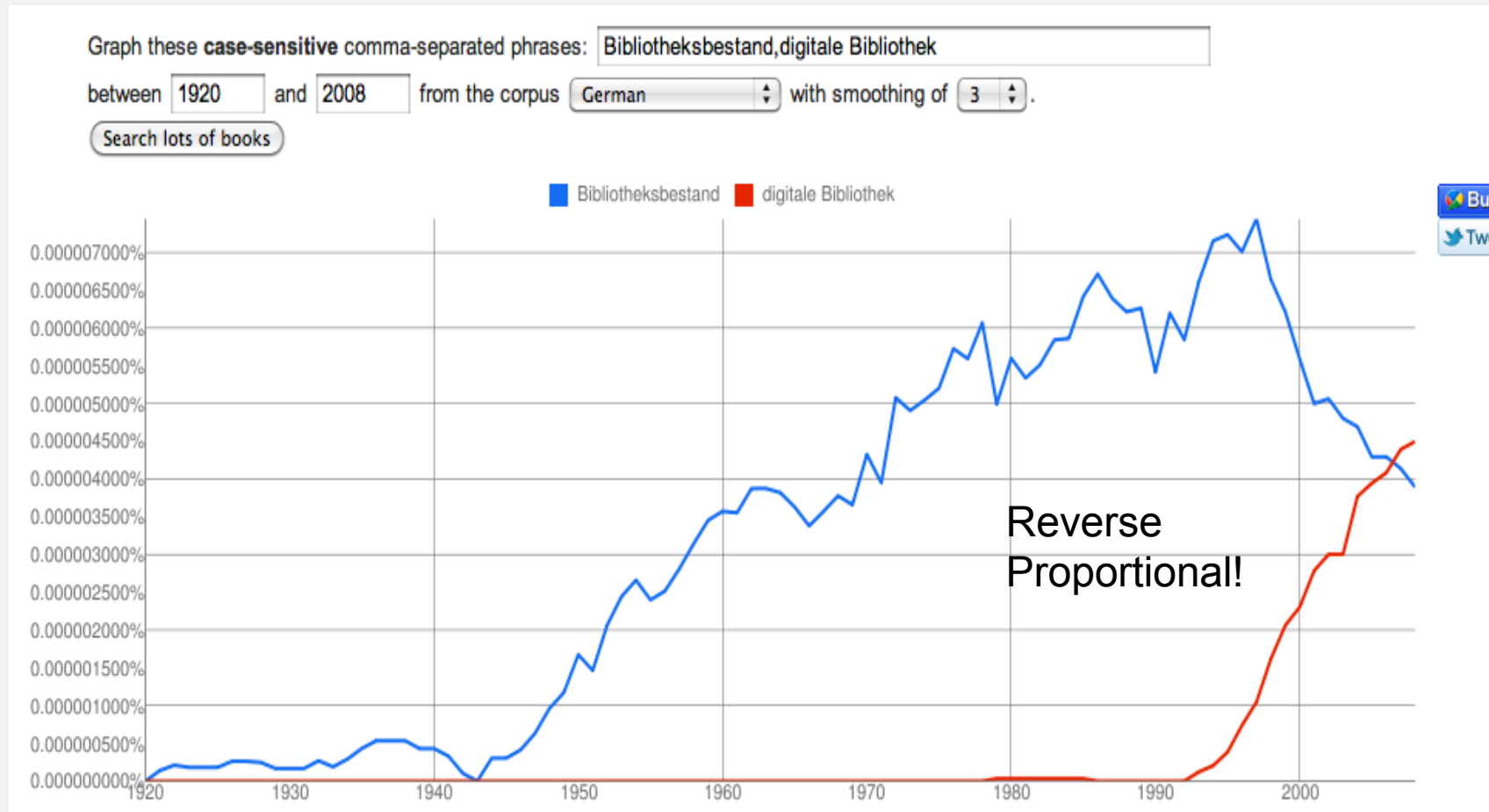
(Shakespeare, Romeo and Juliet (II, ii, 1-2))

Why then do we stick to emptied metaphors?

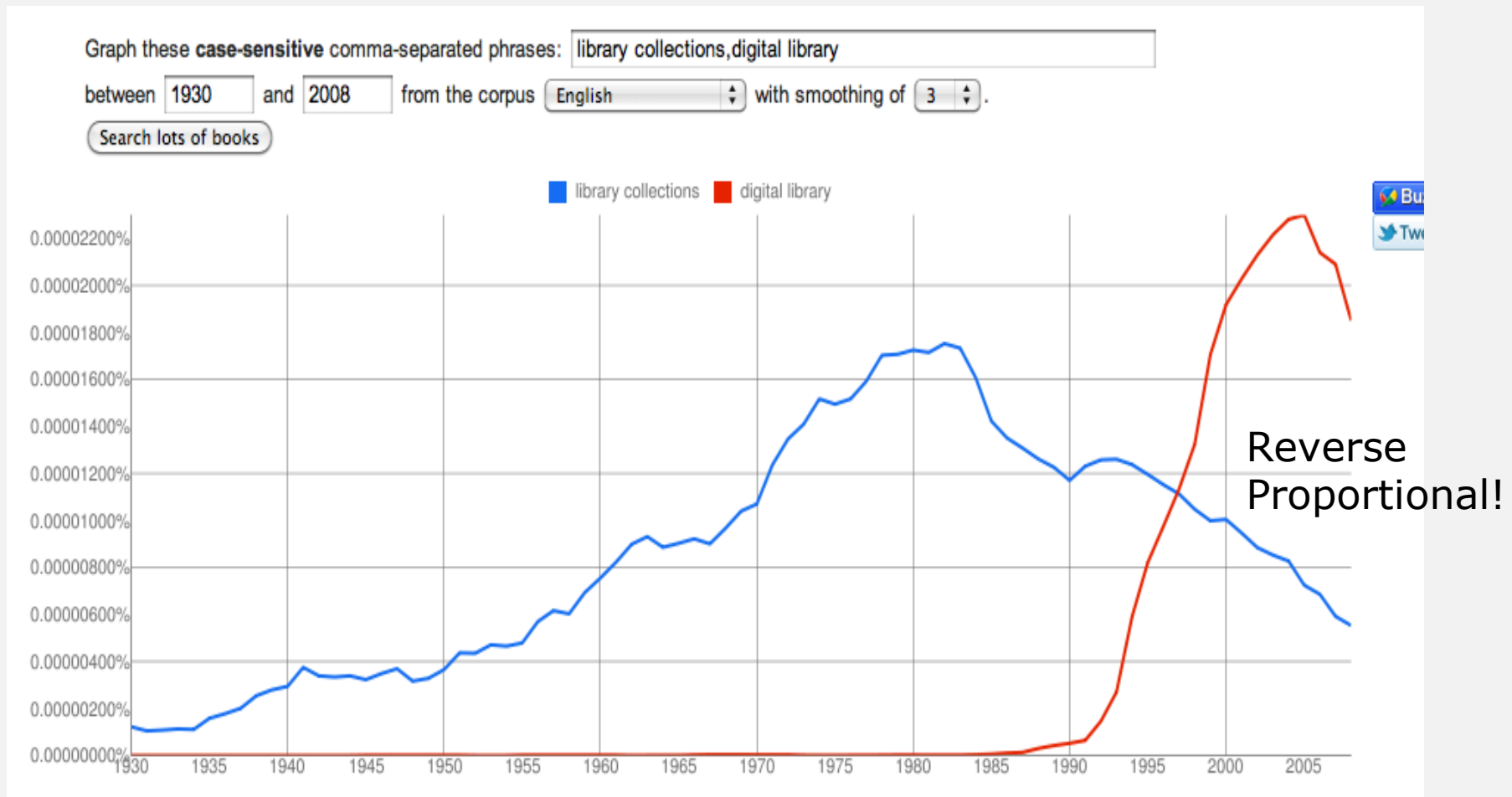
- ... because they **constitute identity** (a very bad reason!)
- ... because they guarantee **institutional persistency** (a fallacy!)
- ... because we are afraid of substantial changes and believe in **things changing only once we use new terms** (dangerously childish!)
- ... or simply because **we do not have new terms** yet?

**Let us then start looking for them!**

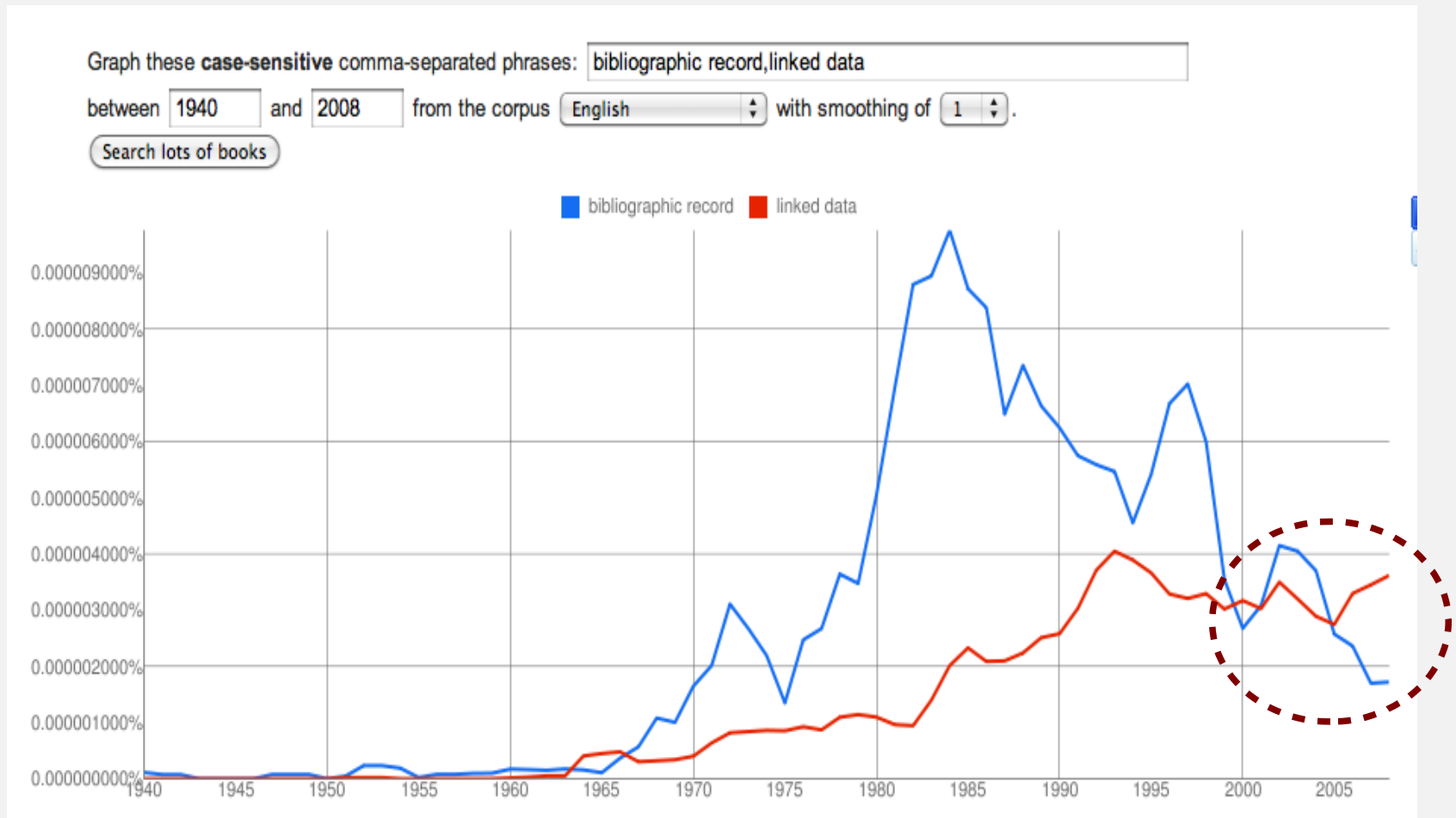
# From 'Catalogues' to 'Graphs': old terms – new terms (1)



# From 'Catalogues' to 'Graphs': old terms – new terms (2)



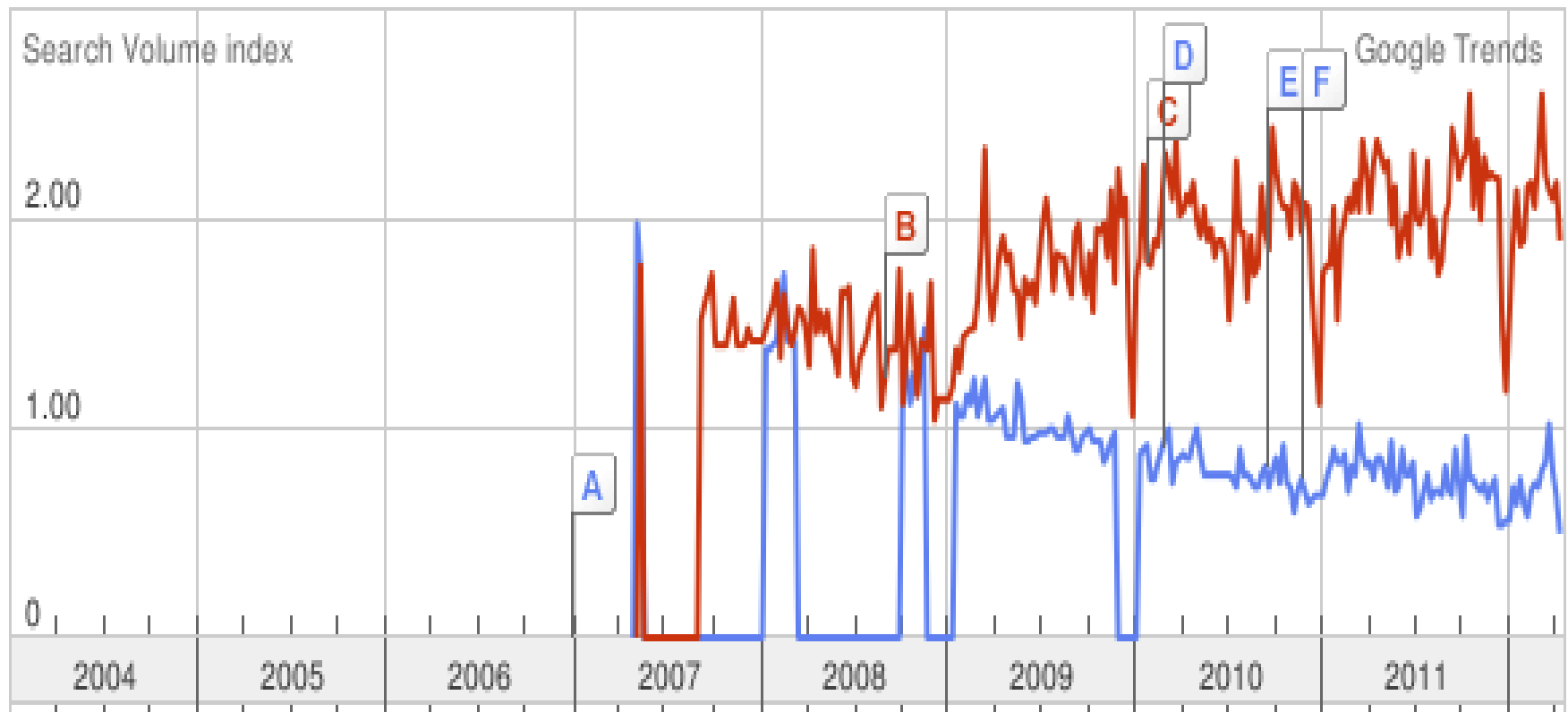
# From 'Catalogues' to 'Graphs': old terms – new terms (3)





# From 'Catalogues' to 'Graphs': old terms – new terms (4)

cataloguing 0.38 linked data 1.00



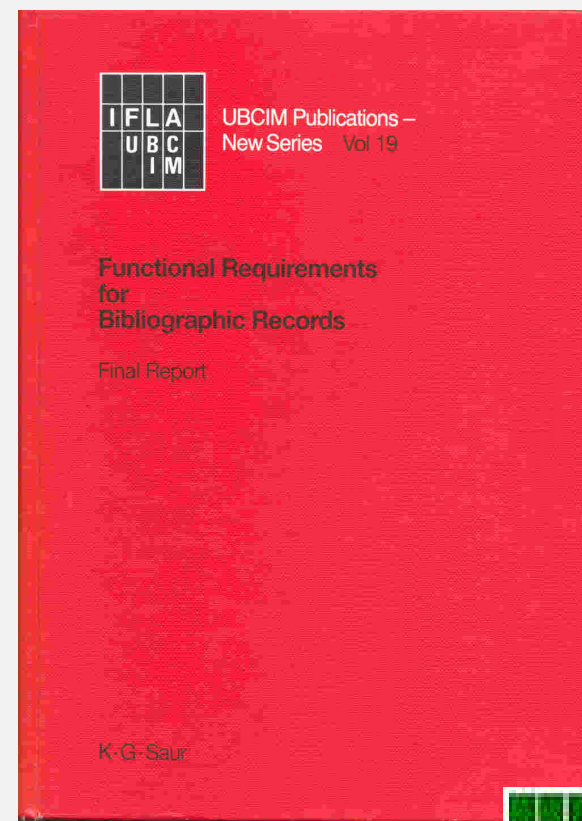
# There is a Life after MARC and AACR2 ...

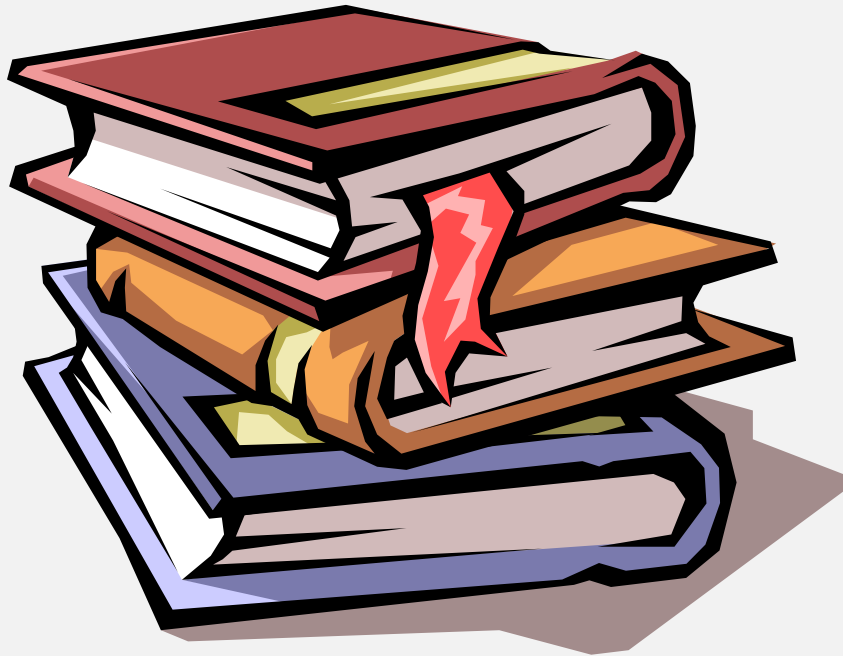
## ... reconceptualising the bibliographic universe

# FRBR

# What is FRBR

- A conceptual model for bibliographic descriptions
- An Entity-Relationship-Model
  - Entities: Groups 1, 2 und 3
  - Relations
  - Attributes (metadata)
- Functional Scope
  - find, identify, select, preserve
- FRBR ≠
  - Data model, data format, rule set, metadata schema, application





- “Book”
  - Physical Object  
(item)
  - 'Publication' in  
Bookstores (ISBN,  
any given item)  
(manifestation)



- “Book”
  - Translator?  
(expression)
  - Author? (work)

# Group 1 – Relations between Entities

Intellectual /  
Artistic Content

Work

Expression

*Is realized through*

Manifestation

*Is embodied in*

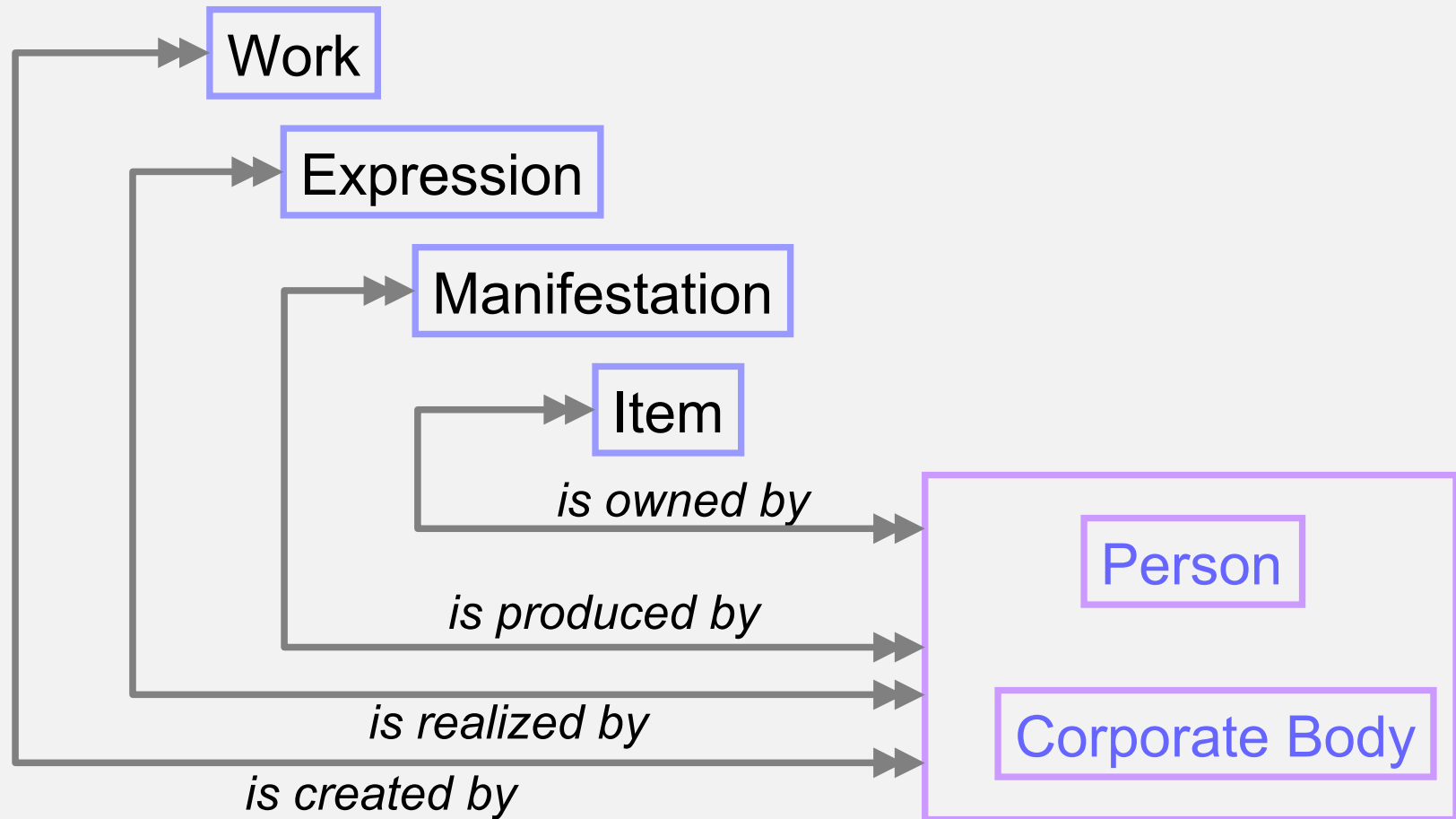
Physical  
Recording  
Of Contents

Item

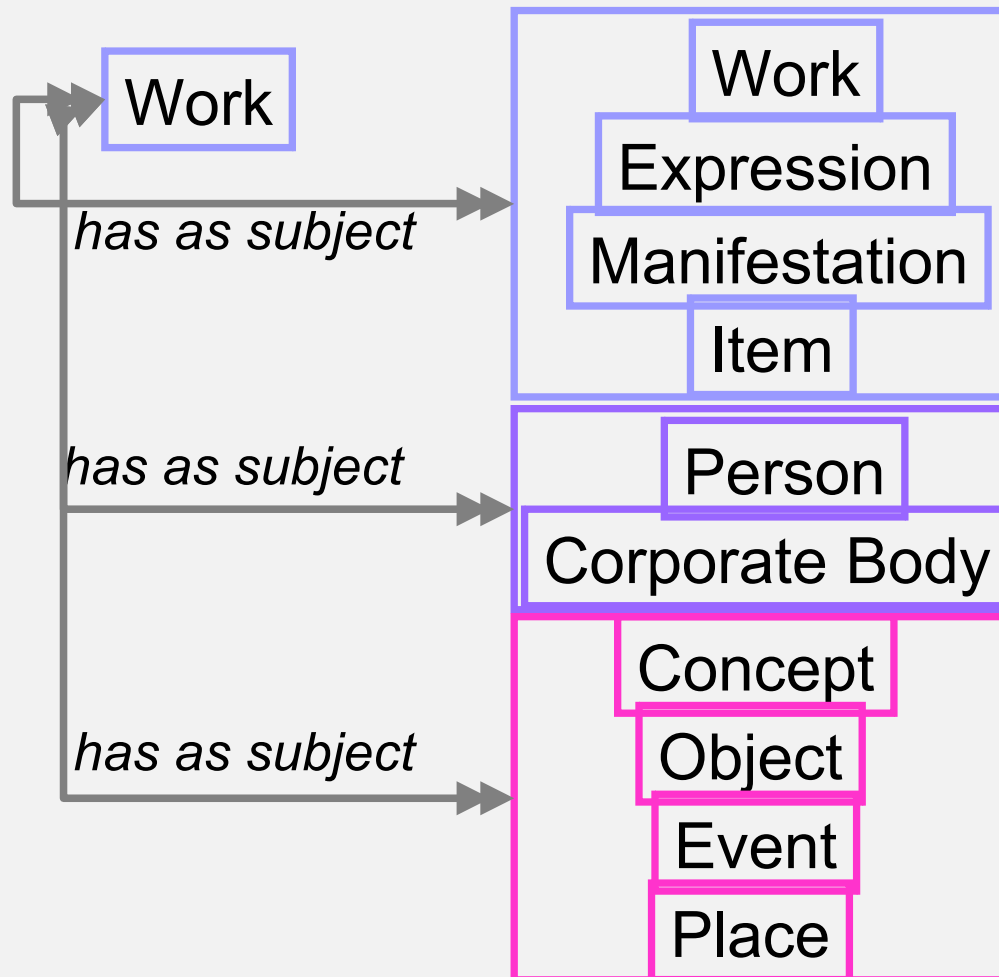
*Is exemplified by*



# Group 2: Relations to persons and corporate bodies



# Group 3: thematic relations of group 1 entities



EQUIVALENT		DERIVATIVE				DESCRIPTIVE	
Microform reproduction		Editions			Free translation		Review
Copy	Simultaneous "publication"		Abridged edition		Summary Abstract Digest		Casebook
		Revision	Illustrated edition		Dramatization Novelization Screenplay Libretto		Criticism
Exact reproduction	Translation		Expurgated edition		Change of genre		Evaluation
Facsimile			Arrangement		Parody Imitations		Annotated edition
Reprint	Variations or versions		Slight modifications		Same style or thematic content		Commentary
					Adaptations		
Original	Same work				New work		

## New Expression

## Cataloging Rules

### cut-off point

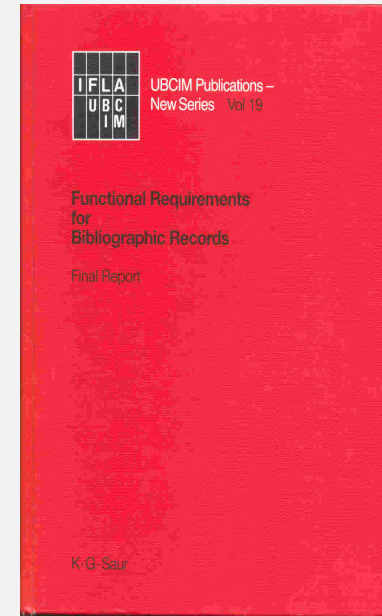
## New Work

B. Tillett  
Dec. 2001

- **Work**
  - ID
  - Title
  - Date
  - etc.
- **Expression**
  - ID
  - Title
  - Form
  - Date
  - Language
  - etc.
- **Manifestation**
  - ID
  - Title
  - Statement of responsibility
  - Edition
  - Statement of Responsibility
  - Carrier properties
  - Access conditions
  - Access modes
  - etc.
- **Item**
  - ID
  - Provenience
  - Location
  - etc.

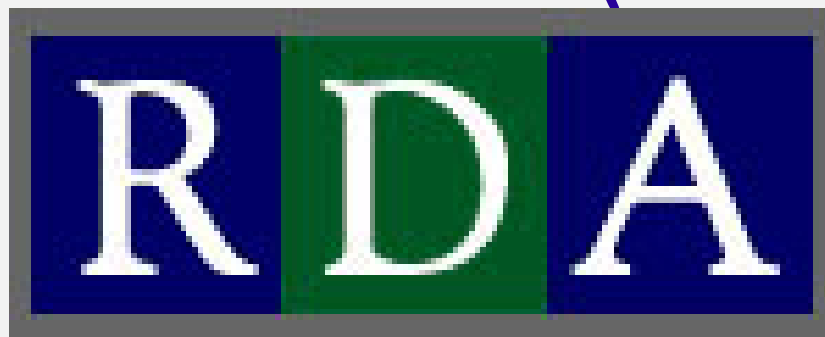
# Application of the FRBR Concept Model

- FRBR is a conceptual model
  - No specific implementation is prescribed
- Implementation examples:
  - OpenVlacc  
[www.bibliotheek.be](http://www.bibliotheek.be)
  - OCLC FictionFinder  
<http://fictionfinder.oclc.org>
  - VTLS Virtua  
<http://bib.uclouvain.be/lib/item?id=chamo:1397047&theme=UCL>



- Better structuring of catalog presentations
- Simplified cataloguing
- Reduced cataloguing load
  - Every work is represented only once for all of its Expressions
  - Every expression is represented only once for all of its manifestations
  - Likewise for manifestations: items just stay what they are.
- Clean layering enables opening of librarian applications to the WWW
- But: ***What is a work???***

# Resource Description and Access (RDA)





# What is RDA?

- Cataloguing rules succeeding AACR2 (originally named AACR3)
  - Established by the AACR Joint Steering Committee
- Clearly going beyond cataloguing rules
- Implements FRBR
- Should be in use from 2013 onwards

# Why not just another Revision of AACR2?



- AACR2
  - 1978
  - 1988
  - 1998
  - 2002

# The Context of 'Cataloguing' has changed profoundly

- Increased range of information carriers ...
- ... and increased depth and complexity of content structures
- Generating metadata isn't a librarian professional privilege anymore
- Increased multitude of metadata formats (DC!)
- Rapidly increasing impact of standards established by W3C

# Changing Terminology

AACR2	RDA
Heading	Access Point
Authority Control	Access Point Control
Authorized Heading	Preferred Access Point
Main Entry	Primary Access Point
Added Entry	Secondary Access Point
Uniform Title	Preferred Title

- Content vs. Presentation of Container Attributes
  - RDA is basically about content, not container
  - RDA-Records can still be displayed in an ISBD compliant way if required.
- Transcription
  - “Take what you see”
  - Error correction
  - Supports automated indexing
- Provides an open door towards the 'Semantic' Linked Open Data Web!
- Is saturated with backward oriented compromises

- Fusion of “MARC must die” and Linked Data @ LOC.gov
- cf. <http://bibframe.org/>
- Bibliographic Framework as a Linked Data Model
- Main Classes
  - Creative Work - a resource reflecting a conceptual essence of the cataloging item.
  - Instance - a resource reflecting an individual, material embodiment of the Work.
  - Authority - a resource reflecting key authority concepts that have defined relationships reflected in the Work and Instance. Examples of Authority Resources include People, Places, Topics, Organizations, etc.
  - Annotation - a resource that decorates other BIBFRAME resources with additional information. Examples of such annotations include Library Holdings information, cover art and reviews.

- W3C, Schema.org
- cf. <http://www.w3.org/community/schemabibex/>
- Based on schema.org:CreativeWork
  - “the most generic kind of creative work, including books, movies, photographs, software programs, etc.”  
(<http://schema.org/CreativeWork>)



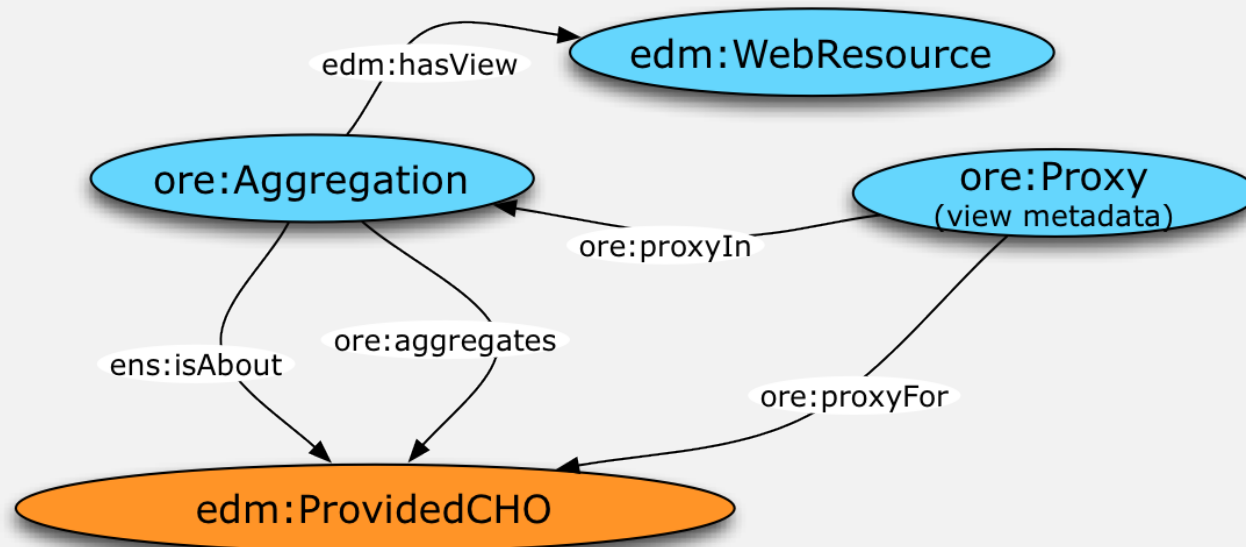
# Ontologically Conceptualizing Bibliographic Entities

- PRISM
  - XML/RDF based industry standard, more at <http://www.idealliance.org/specifications/prism-metadata-initiative>
- BIBO
  - Extends PRISM, more at <http://bibliontology.com/>
- SPAR / FaBIO
  - RDF based representation of FRBR, more at <http://sempublishing.sourceforge.net/> and <http://www.essepuntato.it/lode/http://purl.org/spar/fabio>

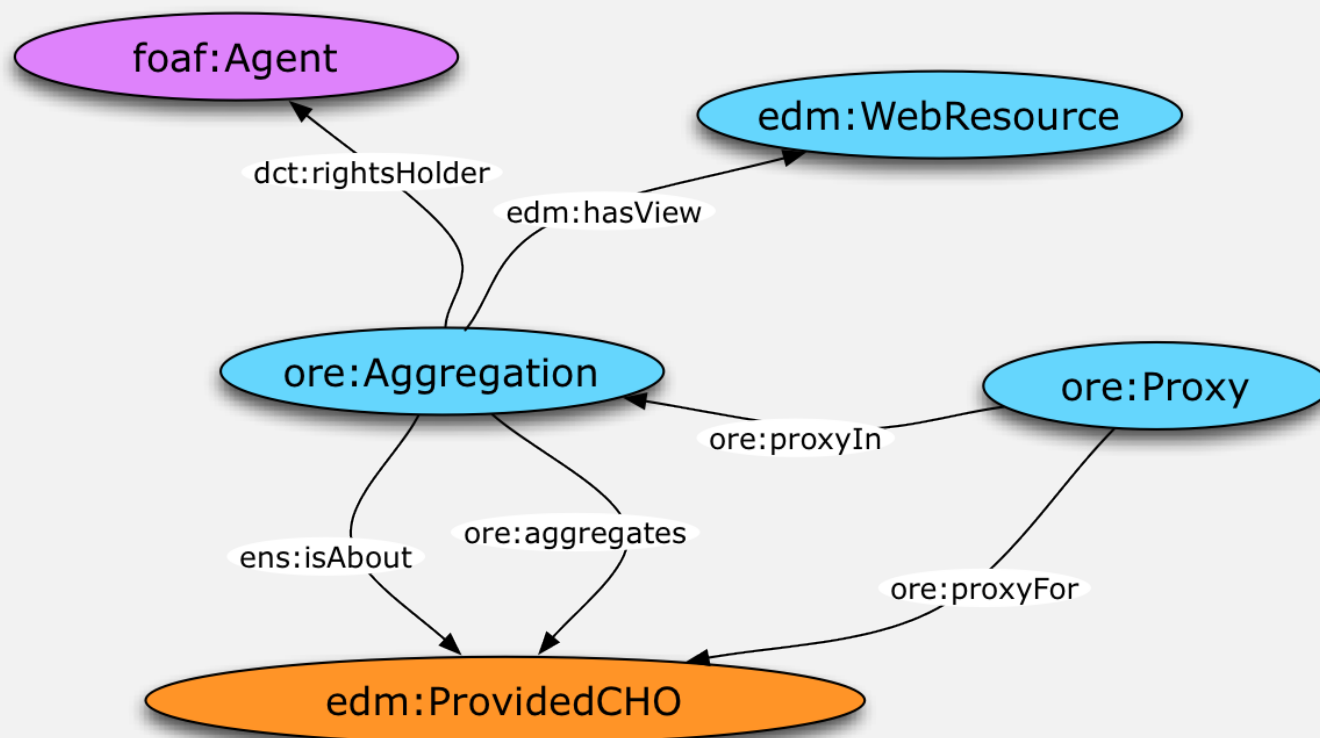


- We have learned some of these lessons in Europeana
  - we dropped the brand “EDL” very early
  - we decided not to have a 'catalogue'
  - We recently consolidated EDM and FRBRoo
- We know that the current portal is not enough
  - we devised the RDF based Europeana Data Model (EDM)
  - we are gradually migrating to EDM based operations
  - we make Europeana part of the Linked Open Data cloud

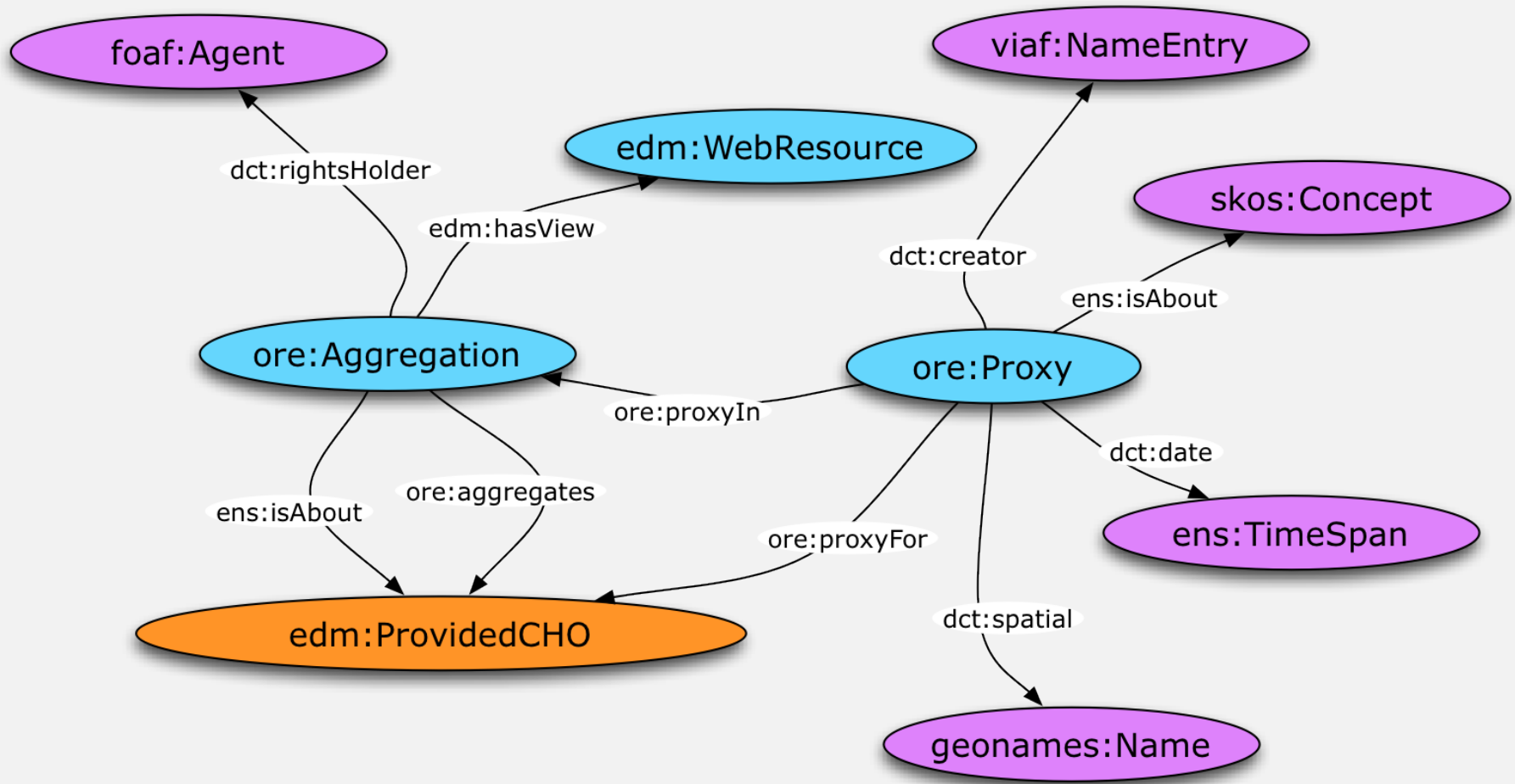
# An Aggregation ...



## ... some context

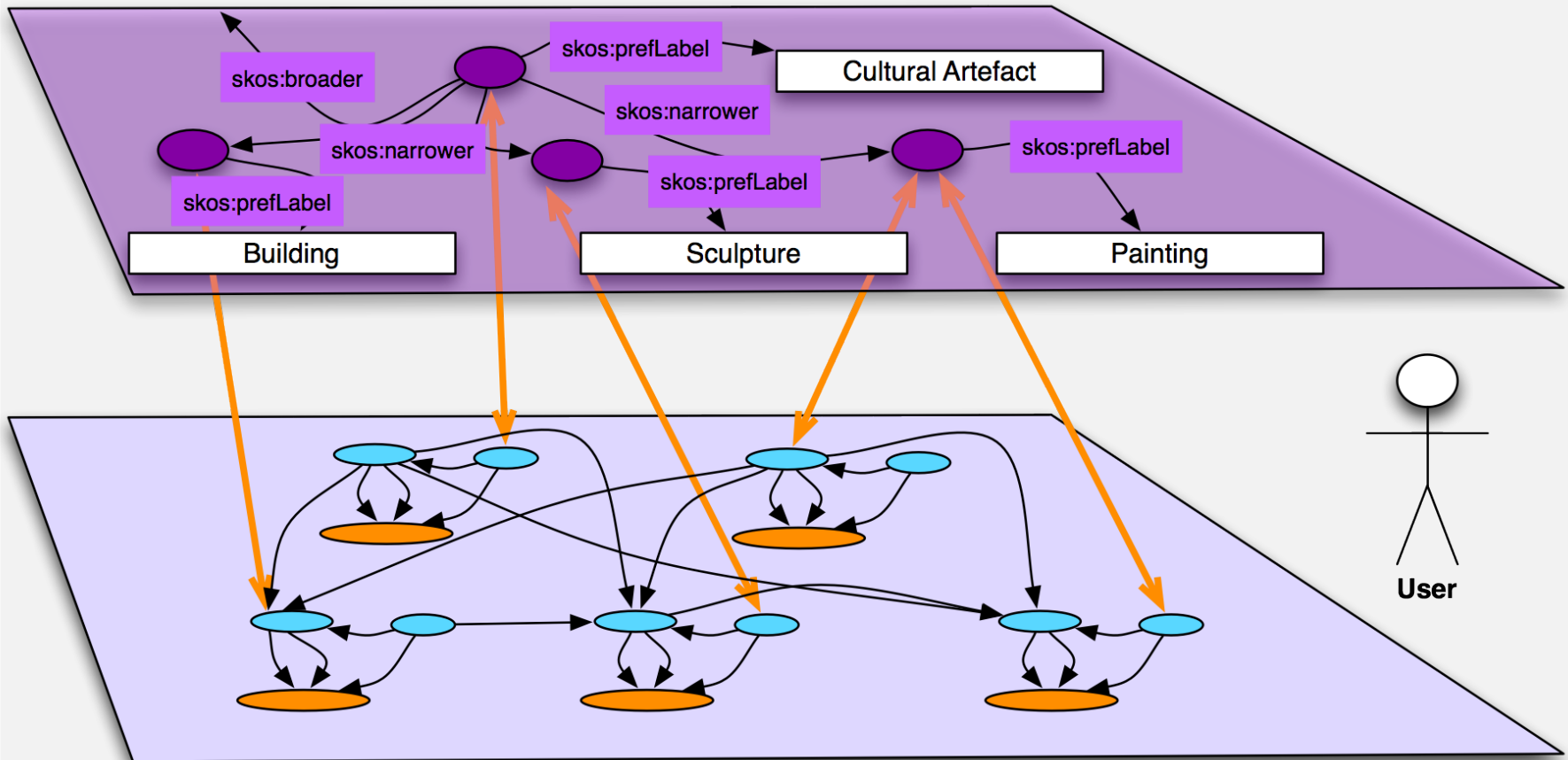


# ... more context



# ... and the Big Picture: Object and Semantic Data Layer

## Semantic Network



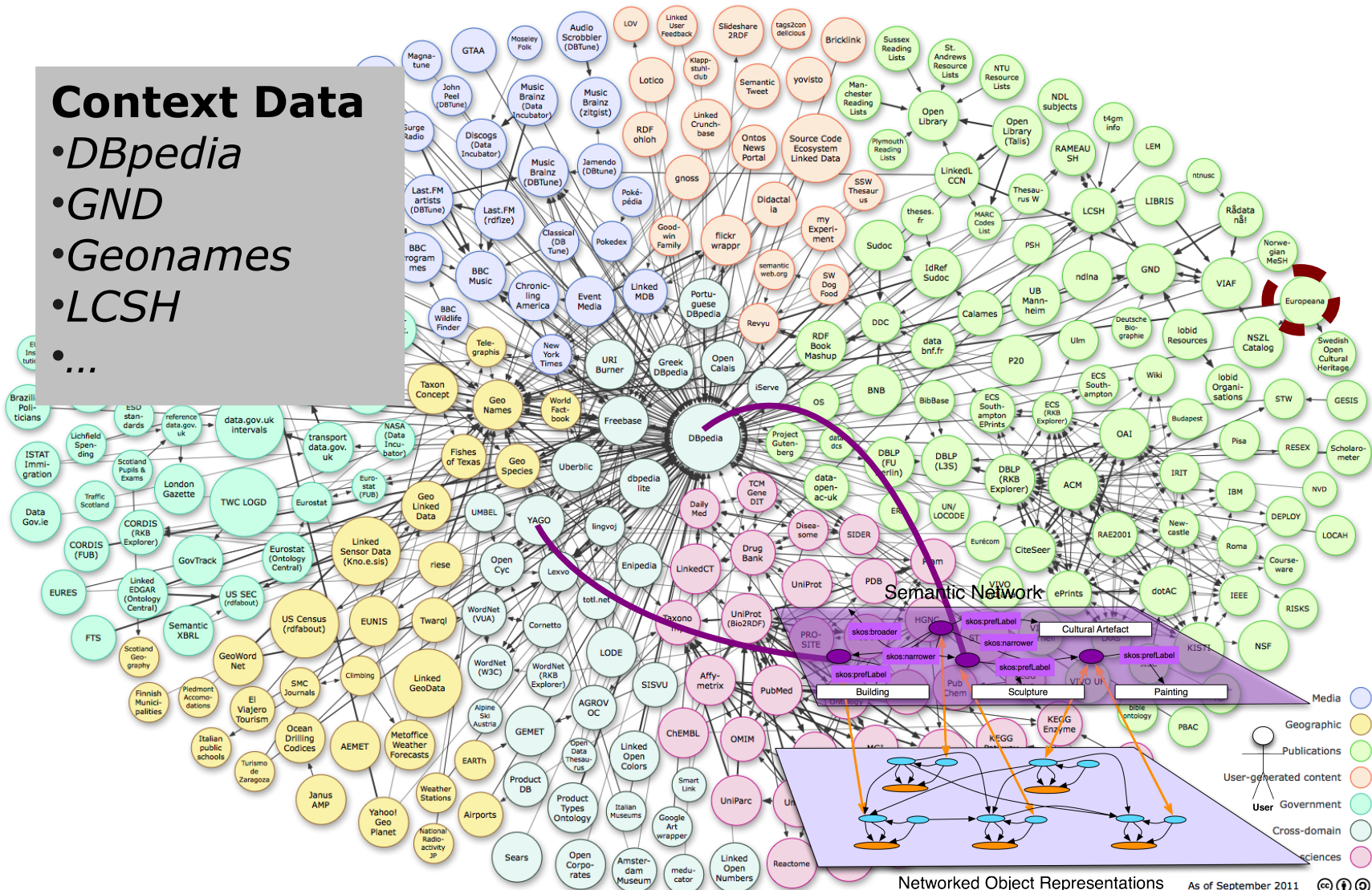
## Networked Object Representations



# EDM and Linked Open Data

## Context Data

- DBpedia
- GND
- Geonames
- LCSH
- ...





# A Triple Win

- **Europeana** wins ...
  - ... as a prime resource for digital humanities scholars
  - ... with a new and clearly focused customer group
- **Digital humanists** win
  - ... with a prime source of corpora on the WWW
  - ... with new research methods based on semantic technologies
- **Librarians** win ...
  - ... a new profile, a new identity
  - ... back in content and knowledge generation
  - ... on a par with scholars again
  - ***All libraries? All librarians??***

# Recap on Librarian Metadata

- Formats
  - πίνακες
  - Catalog cards
  - MARC Records
  - FRBR
  - RDF-Graphs
- Semantics
  - Cataloguing Rules (AACR)
  - RDA
  - Ontologies & Rulesets

# Semantic Annotation: Pundit

**Stefan.gradmann@kuleuven.be**

<http://www.thepund.it/visualization-demos/timeline-demo/>

# Suggested Reading

Gregory Crane (2006): What Do you Do with a Million Books? In: Dlib Magazine, Vol. 12, March. (<http://bit.ly/JhzF90>)

Gutenberg Paranthesis Research Group / University of Southern Denmark: Position Paper (<http://bit.ly/JjGKb6>)

David Parry: Burn the Boats/Books. Presentation to Digital Writing and Research Lab, Austin. (<http://bit.ly/JYLIJIV>)

David Shotton (2009a): Semantic Publishing. The coming revolution in scientific journal publishing. Learned Publishing Volume 22, No 2, 85–94, April 2009; doi:10.1087/2009202

David Shotton et al. (2009b): Adventures in Semantic Publishing: Exemplar Semantic Enhancements of a Research Article (<http://bit.ly/IgT5Km>)

Barend Mons, Jan Velterop: Nano-Publication in the e-science era (<http://bit.ly/IISMGt>)

Alan Renear, Carol Palmer (2009): Strategic Reading, Ontologies and the Future of scientific Publishing. In: Science, August 2009, p. 828 – 832.

## Thank you for your patience and attention

